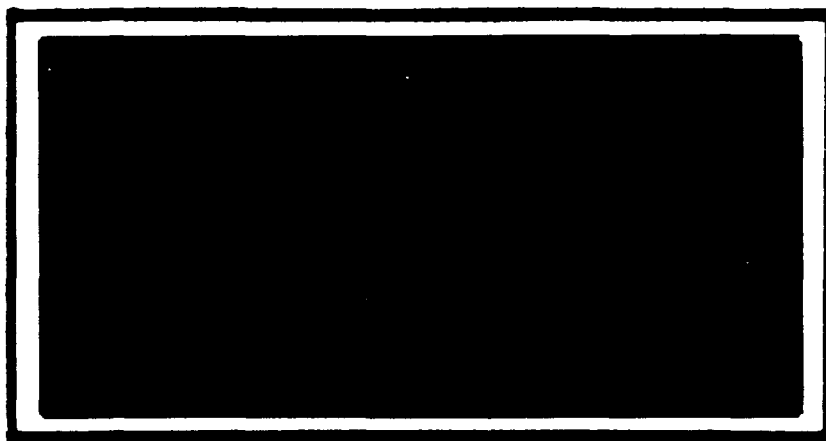


2

AD-A229 425



DTIC  
ELECTE  
DEC 19 1990  
S D D  
CO



EXCLUDED FROM AUTOMATIC A  
DECLASSIFICATION  
EXCLUDED FROM AUTOMATIC A  
DECLASSIFICATION

DEPARTMENT OF THE AIR FORCE  
AIR UNIVERSITY  
**AIR FORCE INSTITUTE OF TECHNOLOGY**

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/LSM/90S-10

DEC 16 1990

AN EXPERT SYSTEM  
FOR THE EVALUATION OF COST MODELS

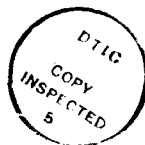
THESIS

Dimitri M. Yallourakis, Captain, USAF

AFIT/GCA/LSM/90S-10

Approved for public release; distribution unlimited

The opinions and conclusions in this paper are those of the author and are not intended to represent the official position of the DOD, USAF, or any other government agency.



Accession For	
NTIS	CPAGI <input checked="" type="checkbox"/>
DTIC	TAD <input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution	
Availability Code	
Dist	Avail and/or Special
A-1	

AFIT/GCA/LSM/90S-10

AN EXPERT SYSTEM  
FOR THE EVALUATION OF COST MODELS

THESIS

Presented to the Faculty of the School of Systems and Logistics  
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the  
Requirements for the Degree of  
Master of Science in Cost Analysis

Dimitri M. Yallourakis, B.S.

Captain, USAF

September 1990

Approved for public release; distribution unlimited

## Preface

The purpose of study was to develop an expert system to evaluate cost models the Air Force uses in estimating the price of future weapon systems. The Air Force has a shortage of cost analysts with experience in the theoretical statistics needed for cost model evaluation. This expert system is a portion of the knowledge of Richard Murphy, Assistant Professor of Cost Analysis at the Air Force institute of Technology(AFIT).

Knowledge acquisition proved to be the most crucial step in programming this expert system. The actual expert system shell used, VP-Expert Version 2.0, is an elementary shell which protracted the process of encoding the knowledge. However, the program successfully models the processes essential to a complete evaluation of a cost model.

In creating the expert system and writing this thesis I have had a great deal of help from several of the AFIT faculty. I am indebted to my faculty advisor, Lt Col James R. Holt, for his continuing motivation and assistance in times of need. Finally, I wish to thank Melina Nicole for her understanding and special support when I was tied to my desk with work.

Dimitri Michael Yallourakis

## Table of Contents

	Page
Preface . . . . .	ii
List of Figures . . . . .	v
Abstract . . . . .	vi
I. Introduction . . . . .	1
The Problem . . . . .	1
A Possible Solution . . . . .	2
Research Goal . . . . .	3
Expected Benefits . . . . .	3
II. Background and Literature Review . . . . .	5
Cost Modeling . . . . .	5
Modeling Problems . . . . .	6
A Literature Review on Expert Systems . . . . .	8
Suitability of Expert Systems . . . . .	8
Expert System Development . . . . .	9
Benefits of Expert Systems . . . . .	10
Drawbacks of expert systems . . . . .	11
Differences in Opinion About Expert Systems . . . . .	12
Unresolved Issues . . . . .	13
Summary . . . . .	14
III. Solution Methodology . . . . .	15
Solution Method . . . . .	15
Problem Definition . . . . .	15
Suitability Of Expert Systems For The Project . . . . .	15
Expert Selection . . . . .	16
Initial Domain Analysis . . . . .	16
Knowledge Acquisition . . . . .	16
Building The Initial Prototype . . . . .	17
Expanding and Verifying the System . . . . .	17
Validate The System . . . . .	19
IV. Results of Research Solution Method . . . . .	20
Following the Method . . . . .	20
The Method . . . . .	20

	Page
Problem Definition . . . . .	20
Suitability Of Expert Systems For The Project . . . . .	20
Expert Selection . . . . .	21
Initial Domain Analysis . . . . .	21
Knowledge Acquisition . . . . .	22
Building The Initial Prototype . . . . .	23
Expanding and Verifying the System . . . . .	25
Validate The System . . . . .	26
V. Conclusions and Further Research . . . . .	28
Further Research . . . . .	29
Appendix A: Sampling of Interview Session Transcript . . . . .	30
Appendix B: Program Code . . . . .	48
Appendix C: Knowledge Summary . . . . .	127
Bibliography . . . . .	138
Vita . . . . .	140

List of Figures

Figure	Page
1. Capture Stage . . . . .	23
2. Example Note Card . . . . .	24



Abstract

The Air Force forecasts the costs of new weapons systems with mathematical cost models. A shortage of cost analysts leads to limited evaluations of the cost models. Frequently, analysts look only at the statistical properties of the regression and not at the logic underlying the theory behind the cost model or equation. This research gathered expert knowledge about the cost model evaluation process and created an expert system of 79 rules to assist analysts in evaluating cost models. The rules were implemented in VP Expert 2.0, an expert system shell. This expert system was verified and validated using the experts opinion with 80% accuracy. The expert system increased evaluation accuracy which leads to an improved weapons system budgeting process.

AN EXPERT SYSTEM  
FOR THE EVALUATION OF COST MODELS

I. Introduction

The Problem

The Air Force uses cost models to estimate the cost of weapon systems. These estimates compose the basis for the Department of Defense budget inputs to the Congress. Cost models that do not predict the cost of future weapon systems accurately directly impact the integrity of inputs to decision makers evaluating management and development issues.

According to Richard Murphy, Assistant Professor of Cost Analysis at the Air Force Institute of Technology, the majority of cost analysts in the Air Force evaluate contracted cost models improperly (15). Evaluation of a linear regression model consists of two steps. First, the theory behind the model form is analyzed. Second, the statistics indicating goodness of fit are either accepted or rejected according to some criteria. Improper evaluation occurs when the first step is overlooked or not addressed and the model is evaluated solely on goodness of fit (15). When models are evaluated improperly or accepted purely on good statistical results, the theory behind the model may be faulty and the model may not be sound. It may not be estimating the costs it was intended to estimate.

Murphy lists several causes of improper evaluation.

1. The number of cost model evaluations needed in the Air Force overwhelms the few qualified cost analysts possessing a strong background in cost modeling theory.

2. Most cost analysts in the Air Force do not have the strong mathematical backgrounds including high level theory courses in statistics which are required to evaluate the reasoning behind the formation of a cost model.

3. A faulty cost model analysis prevents novice cost analysts from learning both steps of the cost model evaluation processes.

4. The shortage of experts provides a limited source of training and guidance for training others. (15)

This paper seeks to solve the root of the problems referenced by Murphy by creating a tool to correctly evaluate cost models.

#### A Possible Solution

Expert Systems are computer programs that model an expert's decision process. An expert system shell captures an expert's logic and knowledge in a computer program. The nature of expert systems provides certain benefits.

1. An expert system can allow an expert to complete a task quicker.

2. An expert system can be an educational tool to the user while it aids in the proper completion of a task.

3. Expert system software can be transported to different

sites so many users can benefit from the knowledge stored in the program (10).

An expert system can solve the problems stated by Murphy associated with cost model evaluation. The expert system program can speed up the cost model evaluation process so that more models can be evaluated properly. The expert system can aid the cost analyst in learning the theory needed for a complete analysis while ensuring proper cost model evaluations. Finally, transportability of expert system software allows the distribution of expert knowledge to many sites providing more sources of training.

#### Research Goal

This thesis develops an expert system to guide a non-expert cost estimator through the evaluation of a cost model. This thesis also evaluates the expert system employing accepted validation and verification techniques.

#### Expected Benefits

The immediate results from a cost model evaluation expert system are quicker, more complete evaluations, broader application of knowledge and techniques, and more educational opportunities. A reduction in the acceptance of inaccurate or invalid cost models should result. The education of cost estimators should eventually expand the base of capable personnel.

In the longer term, the Air Force saves money by avoiding cost models which inaccurately predict the costs of future

weapon systems. This leads to more accurate budgeting. The impact of proper budgetary information passed forward to Congress is better information to make decisions in a constrained budgetary environment. Hence, resources can be used more efficiently at a national budgetary level.

## II. Background and Literature Review

This chapter explains the nature of cost modeling and the problems associated with the evaluation of a cost model. A literature review on artificial intelligence, specifically expert systems, is also addressed in this chapter.

### Cost Modeling

Air Force Systems Command Manual 173-1 defines a cost estimate as the process of projecting financial requirements to accomplish a specified objective. The cost estimate includes "selecting estimating structures, collecting, evaluating, and applying data, choosing and applying estimating methods, and providing full documentation" (1:2-2).

The following example demonstrates the different parts of a cost estimate. Assume that the Air Force needs to estimate the cost to build 100 special engines to be placed in a new aircraft. The cost analyst first selects an estimating structure. This includes setting out the logic which dictates the factors to consider for the cost model. Factors are large categories such as 'technology' or 'complexity' which are seen as important determinants of a change in cost. The analyst then collects all the data available for all known engines and evaluates the data to determine which points will be used for this project. For example, the analyst may find twenty data points or engine types but determines that only twelve of these have characteristics close enough to the new, special

engines to be considered part of the defined population. The analyst then uses statistical theory and other cost modeling techniques to develop and analyze the underlying logic and justification of a cost model.

The analyst considers the factors determined in the beginning of the process and uses linear regression techniques to choose one characteristic or element (called cost drivers) to capture each factor. Different cost drivers are used until the best cost model is found given the data available. An estimating method, such as linear regression, is then used to arrive at a cost model. The final cost model is an equation relating statistically significant characteristics of an existing population in an attempt to estimate the cost of a future population (1:2-2). The cost for the 100 new engines is estimated using this cost model. The entire process and results are documented in a report which becomes the cost estimate.

#### Modeling Problems

The Air Force contracts out most of the cost model development which includes the theoretical and statistical processes described above. These cost models are reviewed and approved by cost analysts for use in estimating future projects.

Undergraduate schools and AFIT short courses teach most cost analysts the mechanics of the statistical processes used in linear regression. The theory dealing with cost modeling

is taught primarily at the graduate level. This theory covers the choice of factors and cost drivers as well as data analysis. There are very few cost analysts with a theoretical education in cost modeling, although almost all have an education in linear regression (15).

There are numerous cost models that need to be evaluated. Because of the shortage of cost analysts with a theoretical background, these models are evaluated by analysts who may not be knowledgeable in the theoretical aspects of cost modeling (15). Although the analyst can look at the model statistics, the analyst may not have the background to evaluate the logic behind the numbers. This can lead to the acceptance of a model that has excellent linear regression statistics but is based on invalid logic or misapplied statistical theory.

The consequences of improper evaluations completed by non-expert cost analysts can be expensive. Murphy theorizes: The Air Force accepts a cost model without a complete evaluation. Suppose the example model does not predict as it is supposed to. Instead, the model provides estimates that are consistently too low. The model is used to develop a cost estimate for 100 engines for a major weapon system. The weapon system is approved and put in the defense budget. The defense budget has a fixed schedule of resources and is completely allocated. Later in the budgetary period, the engine program with the low cost estimate starts to run over cost. This program causes the need for a reallocation of funds, which creates delays. The reallocation leaves another



program short on funding. This cascading effect tips the entire weapons system budget off balance (15).

According to Murphy, one solution exists. The academic instruction offered by the Air Force Institute of Technology Masters Program and other civilian graduate programs provides the cost model theory needed to evaluate cost models properly and completely (15). Increased enrollment of cost analysts in AFIT graduate programs could provide the statistical background needed for cost model evaluation.

It will take time to educate cost analysts in the theory of cost modeling. The expert system is a tool which ensures proper evaluations are completed in the meantime.

#### A Literature Review on Expert Systems

Expert Systems(ESs) improve efficiency and provide consistent quality (7:63). The following paragraphs describe and summarize known facts about ESs as they relate to the cost model evaluation problem.

AI is an attempt to duplicate the human cognitive process with computer programming. S. R. T. Kumara defines an expert system as:

. . . a tool which has the capability to understand problem specific knowledge and use the domain knowledge intelligently to suggest alternate paths of action.  
(11:1107)

Suitability of Expert Systems. Kumara sees the use of an expert system as profitable when:

. . . [1] problems in the domain cannot be well defined analytically, [2] problems can be formulated analytically but the number of alternate solutions is

large . . . , and [3] the domain knowledge is vast and relevant knowledge needs to be used selectively. If a problem fits into any of the above categories it will be worthwhile to construct an expert system. (11:1109)

Kumara's first case fits the situation as presented by Murphy. The domain of this expert system is the theoretical aspects of evaluating cost models. The theoretical aspects of statistics are difficult to define in analytical terms (15).

Expert System Development. ESs use three concepts.

Kumara explains that "expert systems not only use [1] techniques to transfer knowledge but also [2] analytical tools to evaluate it and [3] techniques to learn it" (11:1107).

The cost model evaluation ES may be stored on software and duplicated for use at several sites. This transportability may enable many cost analysts to use the analytical tools captured by the ES to improve evaluations. Kumara sets out the basic parts of ESs as follows:

- (1) knowledge consisting of domain related facts,
  - (2) knowledge consisting of domain related rules for drawing inferences,
  - (3) an interpreter that applies the rules,
  - (4) an ordering mechanism that orders the application of rules,
  - (5) a consistency enforcer, when new knowledge is either created or old knowledge is deleted from the knowledge base, and
  - (6) a justifier that explains the system's reasoning.
- (11:1108)

Michael D. Akers and others are in general agreement (2:31).

Steps one and two above are knowledge acquisition steps.

These steps are accomplished by the knowledge acquisition engineer. The remaining four steps can be accomplished by an expert system shell. Therefore, a student with a basic

understanding of computers can program an expert system if that student can acquire the knowledge from the expert.

Researchers also widely accept the following method used for building an expert system. Captain Jerry L. Moran used these steps as a framework for his thesis (14). Kumara lists the five phases for building an expert system as follows:

- (1) Problem definition;
- (2) Knowledge acquisition, representation and coordination;
- (3) Inference mechanism;
- (4) Implementation: and
- (5) Learning. (11:1108)

The five phases take different amounts of effort with the knowledge acquisition phase being the most time consuming. This development cycle may be used to develop the cost model evaluation ES.

The use of expert system shells helps reduce the time required for system development. Tom Arcidiacono explains "shells simplify and improve transfer of knowledge from a human expert to a ready-made knowledge structure" so knowledge engineers can code their expertise with little help from a programmer (4:56). The use of a shell "dramatically shortens development time . . . [and] serves as a rapid prototyping tool" according to Julie Anderson (3:9).

Benefits of Expert Systems. Expert Systems help reduce the negative effects associated with staff turnover. Beau Sheil states ESs allow "us to capture some part of this otherwise intangible asset [an expert's knowledge] so that it can be preserved in spite of personnel turnover" (22:92).

This quality makes ESs perfect for situations which are characterized by a shortage of experts.

Sheil goes on to explain "since most professionals are expensive, expediting even minor components of their work" could provide significant cost savings (22:92). The use of ESs provides a solution to dependence not only by capturing this knowledge, but by teaching it as well.

Tracing the thought pattern of a knowledgeable cost model analyst is an educational process for the novice. Anderson states most expert systems have "explanation features that trace the steps that the computer follows to reach its conclusion" (3:9). This enables the novice to see the thought pattern of a seasoned cost analyst. Murphy stated one solution to the cost model evaluation problem is education. A cost model evaluation expert system provides another source of education.

Drawbacks of expert systems. ESs do have acknowledged disadvantages which consist of high cost, lengthy development, ignorance of limitations, and misconceptions about AI.

Although expert system shells are inexpensive and can produce a quick system, most companies' needs will extend beyond the capabilities possible with even the most powerful shells. The investment into a large expert system is sizable and the systems may take up to two years to become "fully operational" (12:96). The issue of expanding the cost model evaluation ES is beyond the scope of this thesis.

Another disadvantage of earlier ESs was a lack of

"awareness of its [own] limitations" in situations outside the specified domain (6:44). Sheil presents an illustration of this as follows:

An expert system designed to diagnose heart disease is likely to make intelligent-sounding but completely misguided recommendations for a patient with a broken leg. The danger, of course, is that users will mistake the intelligent tone for real competence and act on the machine's advice. (22:94)

The cause of the danger Sheil describes is a general misconception about ESs.

Misconceptions about ESs come mostly from the nature of the word intelligence when describing AI. Sheil explains, "lacking any precise definition of what it means to be 'intelligent', most people will conclude that an intelligent computer system will behave much as a person would" in similar situations (5:94). The cost model evaluation ES is subject to these misconceptions.

Differences in Opinion About Expert Systems. The experts' beliefs concerning incorporation of AI into commercial uses, tie the different opinions about ESs together. Akers believes ESs are not capable of incorporating behavior variables and so the physical sciences limit their use (2:34). Sheil thinks that misconceptions associated with the "notion of intelligence" limit the use of ESs (22:96). The cost model evaluation ES does not attempt to incorporate behavioral variables or make quantum leaps in intelligence programming. It does capture enough knowledge to be effective at a pointed application.

Unresolved Issues. Researchers need to resolve the issues of programming common sense, programming the ability to learn as experts do, and assigning liability for the decisions of ESs. A brief discussion of these topics follow. However, the ES in this document is not of high enough sophistication or complexity to challenge these areas. The following discussion is included because it contains recurring themes found in the literature during the review.

The problem with programming ESs to have common sense is one of size. Sheil explains it as follows:

Our ordinary interactions assume a great deal of shared knowledge about an enormous variety of topics. But when we judge a task's difficulty, we tend to forget that fact and focus only on the amount of information that must be added to our base of common knowledge. (22:93)

Akers states "to get common sense into the computer AI researchers believe that machines must be able to learn on their own" without program updating (2:34). Dreyfus speculates that both common sense and learning are out of reach for ESs, although researchers are still progressing in such areas (5:45,47).

Another issue is the liability for decisions made by ESs. Arcidiacono states the following:

As more people contribute to the complexity of expert systems and rely on their decisions, responsibility will be diffused. Who will be morally responsible for death resulting from an incorrect medical diagnosis by an expert system: the attending physician who took its advice, the original domain expert, the knowledge engineer, the programmer, or the tool designer? Of seemingly greater practical importance in our society is the question, who is legally responsible? No simple answers exist. (4:56)

Clearly, society must resolve this issue before it allows life entrusting applications of AI.

### Summary

The use of AI and especially ESs is directly applicable to cost model evaluation. ESs can help capture experience and thus aid in teaching cost model evaluation. ESs provide for consistency, quality and improved efficiency and thus upgrade the integrity of accepted cost models. Some disagreements about the use of ESs in life entrusting applications exist but this cost model application does not fall into that category. Agreement exists about the parts of ESs, how to build ESs and the use of expert system shells to start projects.

Unresolved issues include programming of common sense and computer learning. Experiments with these abilities are ongoing. No one knows if common sense or learning can be done within the framework of existing AI technology. Even so, the resolution of these issues does not impact the benefits derived from using ESs.

### III. Solution Methodology

#### Solution Method

This thesis develops an expert system that follows the evaluation pattern of an experienced cost estimator. The system will contain rules reflecting the knowledge attainable from written literature incorporated with rules drawn from the expert. The system will then be validated by the expert.

This section defines the steps for the development of the expert system. This list is a conglomeration of several lists described in available literature by Anderson, Freiling, Hazen, Kumara, McMain, and Simmons.

Problem Definition. The definition of the problem is the first step in finding the solution (11:1108). The specific problem is to develop a system that will allow the user to evaluate cost models on the basis of sound theory and logic.

Suitability Of Expert Systems For The Project. Hazen offers a twenty-three question guideline to evaluate the applicability of expert systems to a problem (9:32). Hazen breaks these down into five basic categories of concern:

- I. Realistic
    - Volatility and recurrence of problem
  - II. Justified
    - Importance of improving problem solving capability
  - III. Expertise
    - Availability of experts
  - IV. Task
    - Applicable size and nature of problem
  - V. Other
    - Cognitive skills needed to solve the problem
- (9:35)



The problem is evaluated using these criteria. The process takes place in the presence of the expert so that a clear understanding is formed concerning the nature of the problem.

Expert Selection. This expert system is modeled after Richard Murphy, an Assistant Professor of Cost Analysis. As part of the Air Force Institute Of Technology faculty, he is available, knowledgeable in the domain of cost analysis, and posses several years experience.

Initial Domain Analysis. The initial domain analysis allows the knowledge engineer to gain a basic familiarity with the problem (13:65). In the area of cost analysis, learning the material covered by the AFIT classes offered in the PCE program as well as the masters program constitute the written knowledge. These techniques include an understanding of the analysis of variances or ANOVA tables and tests for statistical problems such as collinearity. The knowledge engineer will have completed these courses before the expert system construction is begun. Freiling calls the stage the familiarization process (8:42).

Knowledge Acquisition. This project uses interviews as the basic process for knowledge acquisition. The sessions are tape recorded and the knowledge engineer takes notes that apply to the organization of the interview. Simmons suggests that the contents of the interview are then transcribed, and reviewed by expert to verify the accuracy of the session (23:163). Alternate means of knowledge acquisition include a critique and feedback on accomplished rules as well as

spontaneous memos or notes.

Building The Initial Prototype. Simmons breaks initial prototyping down into two stages. the capture stage and the organization stage (23:163).

The capture stage refers to the process of documenting the objects, relations, and actions that make up the knowledge (3:163). This step calls for the identification of the basic issues. The organization process refers to ordering the knowledge in such a form that it is ready for mapping to rules (3:163). This step is a refinement of the captured knowledge. The knowledge engineer concentrates on the applicability of different inference strategy designs at this point. The VP-Expert expert system shell will be used for this project. Using this shell, a control strategy is now formulated and coded. Rules are added and the control mechanism is constantly updated until the project reaches a working state. At this point, some attention is given to developing a preliminary interface design and it is introduced into the program code. Freiling suggests a development structure similar to that above in regards to the sequencing of inference control and interface design (8:43).

Expanding and Verifying the System. As the knowledge acquisition process continues, the knowledge engineer continues to add rules and check the control mechanisms for logical and efficient construction. Once the system is running, verification of coded rules is accomplished. The expert confirms the truth of the rules, while the knowledge

engineer analyzes the rules for other verification issues. Tin A. Nguyen breaks verification down into two sections: consistency and completeness (20:69).

Consistency of the system is assessed with five inspections. Any conflicts revealed by these tests are corrected by the knowledge engineer. The first is an inspection for redundant rules. Nguyen defines rules as redundant "if they succeed in the same situation and have the same conclusion" (20:71). The second test is for conflicting rules. Nguyen defines two rules as conflicting "if they succeed in the same situation with different conclusions" (20:71). The third inspection is for subsumed rules. Nguyen explains "one rule is subsumed by another if the two rules have the same conclusions, but one contains additional constraints on the situations in which it will succeed" (20:71). The fourth test is for unnecessary IF conditions. Nguyen puts forth the following explanation:

Two rules contain unnecessary IF conditions if the rules have the same conclusions, an IF condition in one rule is in conflict with an IF condition in the other rule, and all other IF conditions in the two rules are equivalent. (20:72)

The last inspection for consistency is for circular rules which are a set of rules that form a cycle (20:72).

Completeness of the system is assessed with four inspections. These inspections look for omitted or missing rules. If this situation occurs, the appropriate rules are added. The first inspection is for unreferenced attribute values. This happens when all the "legal values in the set

are covered only partially or not at all" (20:73). The second inspection is for illegal attribute values. This is simply when a rule refers to an attribute value that is not in the set of legal values (20:73).

The third inspection is for unreachable conclusions.

Nguyen offers the following explanation:

The conclusion of a rule should either match a goal or match an IF condition of another rule (in the same rule set). If there are no matches for the conclusion, it is unreachable. (20:73)

The fourth inspection is for dead end IF conditions. Nguyen offers the following explanation:

To achieve a goal . . . , either the attributes of the goal must be askable (user provides the information), or the goal must be matched by a conclusion of one of the rules in the rule sets applying to the goal. If neither of these requirements is satisfied, then the goal cannot be achieved (that is, it is a dead\_end goal). (20:74)

Validate The System. The system is validated using a predictive validation process. Robert M. O'Keefe describes the predictive test as follows:

An expert system is driven by past input data from test cases, and its results are compared with corresponding results - either known results or those obtained from the human expert. (21:96)

The results in this case are evaluated by the human expert. The expert will classify the system at percent ideal, percent acceptable, percent suboptimal, and percent unacceptable (21:92).

#### IV. Results of Research Solution Method

##### Following the Method

This section follows the format set out in chapter three and records the results, problems and modifications.

The Method. The methodology listed in chapter three started with the problem definition followed by an analysis of the project for suitability of expert systems. These steps were followed by expert selection, initial domain analysis, knowledge acquisition, initial prototyping, expansion and verification, validation, and evaluation.

Problem Definition. The specific problem was to develop a system that allows the user to evaluate cost models on the basis of sound theory and logic as opposed to the strength of the linear regression results. This problem statement was arrived at late in the process. Even though the expert expressed the problem early in the process, it was difficult to interpret it clearly.

Suitability Of Expert Systems For The Project. The problem meets Hazen's five criteria explained in chapter three for the suitability of expert systems. The first criteria is realism which addresses the volatility and recurrence of problem. The nature of the problem of improper cost model evaluations is identified by Murphy as a constant and recurring problem (15).

The second criteria is justification which deals with the importance of improving problem solving capabilities for

this area. Murphy addressed this topic in the personal interview in March 1990. There is a shortage of cost analysts who possess an understanding of the theory behind the statistics needed to properly evaluate a cost model (15). Therefore, an improvement in problem solving capability will enable more cost models to be evaluated correctly.

The third criteria is the availability of expertise without which it would be difficult to obtain the knowledge to build a system. The expert for this system resided at AFIT which made it easy to schedule knowledge acquisition sessions.

The fourth criteria is the size and nature of the problem. In this case, the scope of the problem was small enough to handle with the available time and computer resources. The nature of the problem is characterized by a lack of education in certain statistical areas. Education is one advantage to expert systems in that the use of the system can provide knowledge the novice can benefit from.

The last criteria is the presence of the cognitive skills needed to solve the problem. The knowledge engineer is proficient with the use of VP-Expert and the nature of expert system

Expert Selection. Richard Murphy, an Assistant Professor of Cost Analysis is the expert. Choosing the expert was a function of availability as well as knowledge.

Initial Domain Analysis. The initial domain analysis allows the knowledge engineer to gain a basic familiarity with the problem (13:65). The domain for this problem consisted of the theory of statistics as well as an

understanding of the Air Force cost estimating procedure. The knowledge engineer familiarized himself with statistics by completing four graduate math courses at AFIT. They were QMGT 670 Statistics for Cost Analysts, QMGT 671 Statistics for Defense Cost Modeling, QMGT 672 Model Diagnostics, and QMGT 673 Cost Estimating for Weapon Systems Production. The techniques taught in these courses include a basic understanding of statistical distributions, an in depth analysis of variances, model building techniques, and several statistical tests for theoretical soundness. The knowledge engineer used the Air Force Systems Command Cost Estimating Handbook to better understand Air Force cost estimating procedures. This was a large amount of knowledge to assimilate and a large amount of time was required to take the courses.

Knowledge Acquisition. This project used interviews as the basic process for knowledge acquisition. Two preliminary sessions took place on 9 April 1990 and 30 April 1990. The knowledge engineer taped the sessions and took notes that applied to the organization of the interview such as diagrams drawn on the board. The contents of the interview were then interpreted and charted on poster board. These poster boards were reviewed by the expert to verify the accuracy of the sessions (23:163). The knowledge representation is important to a successful system. When spacial relationships are established, it becomes much easier to partition the information into manageable parts. Without

this visual linking, a firm conceptual grasp of the variable inter-relationships would have been extremely difficult.

Building The Initial Prototype. This step calls for the identification of the basic issues. The knowledge engineer and the expert agreed on five major divisions of the knowledge.

After the mapping of variables, the knowledge was broken down into many parts. The first portion of the knowledge contained considerations for the population and the definition of the cost model. The second portion dealt with data analysis. The third portion of the knowledge captures the issues dealing with the identification of the model variables. The fourth portion of the knowledge considers the specification of the variables or what mathematical form they take on. The fifth portion of the data deals with the analysis of the model statistics.

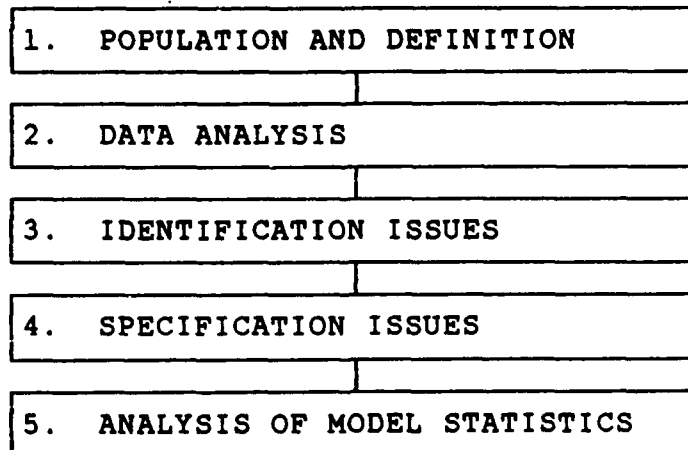


Figure 1. Capture Stage



The knowledge engineer then physically mapped all the knowledge on large sections of poster board and connected dependent portions with lines.

The organization stage refers to ordering the knowledge in such a form that it is ready for mapping to rules (3:163). This step is a refinement of the knowledge gathered in the capture stage. This step transferred the knowledge charted on the poster board to individual three by five inch note cards. Each note card was labeled with one of the five major categories identified in the capture stage described above.

The knowledge engineer further divided the note cards into smaller categories which became the basis for the control strategy. The actual program was then formulated and coded. Rules were added and the control mechanism was constantly updated until the project reached a working state.

POPULATION AND DEFINITION	
—The cost model should contain a list of —	—
—cost drivers that were considered for —	—
—the equation. If not, the cost model may—	—
—not be capturing some of the variation in—	—
—cost—	—

Figure 2. Example Note Card

During this process, significant problems with the software were overcome. The major problem was the inability of VP Expert to handle large applications efficiently.

Although VP Expert was easy to learn, batch files had to be used to sequence programs and this led to slower run times. The completed program contained ten programming modules and one database. Most of the rules at this point dealt with acquiring user responses and manipulating the database. The system acted more like a computerized checklist. This was due to the way the expert presented the knowledge to the knowledge engineer.

It is important to note here the expert system stores knowledge in two places. The first is the rules presented in the VP-Expert program modules. The second is the database structure and relationships. Although the expert system was running, most of the expertise was contained in the database at this point.

Expanding and Verifying the System. The verification of coded rules was accomplished by inspection of the knowledge engineer. The rules were relatively simple so little problems occurred. The database was also reviewed and the fields and relationships were redefined. After the first transition, the expert also reviewed the rules. Although previous interviews were captured on tape and transcribed, only about 50% of the knowledge was accurately captured according to the expert (19).

The knowledge engineer did nothing wrong but still only captured half of the data. This is attributed to several factors. First, the expert had given an organizational structure in the initial interviews which he later altered

when evaluating the systems. Since the programming modules were structured around initial outline, the expert viewed the program as out of sequence. Second, the expert added rules and conditions during the interview that were not expressed during initial knowledge acquisition sessions. This may be an indication that the expert did not realize all the rules needed to completely evaluate the situation. Only when he started to evaluate the prototype did he realize there were more rules needed to complete the system. This review substantiated the concept of iterative knowledge development.

The programming modules were modified and several rules were added and others rewritten. On the second review, the expert found about 80% of the knowledge was accurately represented at this point. This is a 30% improvement over the initial verification interviews.

The final domain knowledge summary corrected grammatical errors and is available in appendix C. Other cost analysts should review this knowledge. They may agree or disagree, add or delete rules. This process develops and matures the skills of cost analysts and the body of domain knowledge. This is a significant contribution to cost modeling. This career field now has a checklist to evaluate cost models.

Validate The System. Proposed validation included face and predictive tests. This was dependent on the available cost model cases to use for the tests.

The model experienced some theoretical problems when the validation tests were actually explored. This program matches

the knowledge of the model against the knowledge contained in the expert database. The program itself turned out to be a computerized checklist which interacted with a database developed by experts. This program is extremely successful in accomplishing this interaction.

Although the program contains knowledge, the database associated with it provides the expertise in each specific field, such as engines. The expert database used for the purposes of testing the model was not developed by any scientific method. It was pieced together simply to allow the VP-Expert program to process the checklist.

A validated database for aircraft, for example, would contain the factors associated with engines as well as the cost drivers that capture those factors. The database would also contain the specific form of the cost drivers and how the cost drivers behave with respect to cost.

The situation described above allowed the model to be validated only by the opinion of the expert. Although the expert found the program a useful tool as a checklist, the program could not be used to evaluate cost model documentation until the database dealing with the specific area is developed and validated.

## V. Conclusions and Further Research

Expert systems are not all they are cracked up to be. It seems once you understand how the systems are built, there is little mystery as to how the expertise is accomplished. This system is an extensive checklist with a fair amount of database manipulation. The most useful knowledge obtained through this process is the list of items needed for a proper Statement of Work. The model itself turned out to be highly dependent on the database associated with it. The database contains large amounts of knowledge primarily in the interrelationships of the knowledge. Validation became just a review of the database.

The benefits of the extensive checklist can be summarized as follows.

1. The checklist accumulated knowledge never before gathered in this format.
2. The checklist organized the data to show the method of solving problems associated with cost model analysis.
3. The checklist identifies missing information and further needs for proper cost model evaluations.
4. The checklist shows the next step in the process giving the analyst a systematic method for evaluation of cost models.

What we really need is an accumulation of all the knowledge in this area. With a complete knowledge base a more

powerful shell could be to apply deductive and abductive reasoning that captures the expert's skills.

Further Research This expert system provides the control structure for evaluating cost models. With some adjustment, the systems could be modified to help the user create a cost model. This would be a very useful system and it appears the need exists judging from comments made by Murphy. Other follow-on efforts furthering this system would be the review and expansion of the checklist and the development and validation of a specific system database. Although this expert system interacts with an engine database, that database is not validated.

## Appendix A: Sampling of Interview Session Transcripts

NOTE: This appendix is a transcript of the initial knowledge acquisition interviews. Assistant Professor Richard Murphy is being interviewed. The interviewer's comments and questions are capitalized. Any unclear portions of the taped transcript are indicated with the following notation: (...).

WHAT IS THE PROBLEM CONCERNING COST ANALYSIS?

Contractors, who we are getting paid big money, are going out and developing cost models and then coming in with these studies and proposing these models to us to use to develop our own cost estimates. We don't have the capability or the skills to evaluate the quality of the research that they've done. So a lot of times we are paying off on models, and then six months or a year after we've paid for them, we find out they are no damn good when someone who has the kind of expertise to really evaluate them gets a chance to see them and then says "Hey this thing is all full of holes, It's worthless" and of course you have already paid for it so what do you do.

HOW MANY EXPERTS ARE THERE? IF IT WAS A PYRAMID WITH THE MOST EXPERIENCED AT THE TOP, HOW FAST DO WE COME DOWN THAT PYRAMID?

There are very few people who have the capability of really evaluating the research that goes into developing a cost model. Part of the problem is that they don't develop that

expertise because we don't do it within the Air Force. We almost always contract out the development of the cost model. So what happens is because the contractors are doing almost all their model development for us, we are using models to develop estimates, but we don't develop our in-house expertise in terms of how to develop cost models because we always contract them out. If you would develop some models on your own, you would develop that expertise, and then you would be in a position where you could evaluate the research being done by other people. But, we do very little in house model development and a lot of times the model development we do is done on a very superficial level because it is being done by people on a kind of sporadic or as required basis and they don't develop the expertise that way it takes a little bit of time to develop time good expertise in cost models.

#### WHERE ELSE COULD YOU LEARN COST MODELING?

1. GCA
2. PCE 550 but that is concentrated and in a short period of time
3. Locals can go to the school at night

#### WHAT ARE THE CONSEQUENCES OF A BAD EVALUATION OF A COST MODEL?

1. Bad cost estimate
2. Reprogramming action causes a lot of perturbations
3. Program ends up costing more than it should have in the first place.

Other consequence of course is that by underestimating program cost we allocate the resources we have available among other



programs and then when some program comes up short that doesn't effect just that program it affects the whole structure of programs because we end up having to reallocate between programs and all that type of things. So that cost impacts multiple impacts instead of just one. The ultimate impact is that we end up paying a lot more for a lot of things so it is wasted resources.

I tend to think more in terms of how I would go about developing a cost model, because when I am evaluating one I basically go through the same process, even though sometimes I will accept certain things that the developer has done. I evaluated one model where he looked at (...) models. I looked at all the new models and didn't look at any other transformations because I accepted that was the way he wanted the structure modeled. When you are starting out probably the first thing you are going to run into is (...) you've got to look at all three of them simultaneously. That's the idea. What are the significant cost drivers which should be considering (...). Well, the availability in general, because you may (...) you think there is some significant cost driver that should have been included that they didn't include, the question is 'why not?'. If they didn't include it because there was no more data available, then that is just a limitation you are going to have to accept. If there is data available and they didn't include it then that becomes a deficiency in their model that probably should have been addressed. Of course, (...) it depends on how you define your

population in that particular issue. One of the things you deal with when you are evaluating a model like this, depends on whether your evaluation is general or specific. If you are evaluating a model that you want to see what applies to your particular estimate, that makes it a lot easier because you've got the criteria, the relevant (...) data and so on, if that does apply to your particular problem. If you evaluate a model that you've asked the contractor to develop - just a general model that you're going to make available to the community to use - but you're not sure how people are going to use it, then your relevant range is no longer defined by the problem you're looking at, but by the availability of the data. You can more or less accept that and just make sure that it gets documented. Sometimes if you have knowledge in a particular situation range, you may want to restrict the range of your data. (...) You may want to include those estimates you have here, but if you estimate them here somewhere, you may want an expert for this data.

WHAT DO YOU GET OR RECEIVE WITH MODELS? DOES THE MODEL CONTRACTOR USUALLY KNOW WHAT WE ARE SHOOTING AT TO USE IT FOR?

No. Generally speaking, what I have seen is that the documentation of contractors is usually lousy and the reason for this is because they give us exactly what we ask for. We do not wait for good statements of work to go with the RFP (...) We say we will investigate this problem. We don't put any requirements in terms of documentation or in terms of types of analysis that we want them to do. There is a lot of

lack of structure there and they pretty much give us what we ask for.

These three kind of have to be looked at all at one time on developing the kind of approach that it would take. If somebody gives me a model. . .

SAY IT'S LIKE A TREE - TO DECIDE WHAT BRANCH YOU GO DOWN, DO YOU FIRST LOOK AND SAY, HOW MUCH (...) DO THEY GIVE ME, WHAT DATA POINTS DO THEY GIVE ME, WHAT (...) THEY ARE USING, AND USE THAT AS A CRITERIA TO GO OUT IN MORE DIRECTIONS...

The first thing I would look at is that I would ask the question 'why is the model being developed?' There are two aspects: one is 'what is the development cost?' and then you have to know something about the research and development cost. Then you have to look at something about the definition of the system. That starts giving you some idea of the population to go to. When somebody says develop a model for helicopter turbine engines, then the question that immediately pops into my mind is "Are we talking about helicopter turbine engines for performance for the helicopter for the past 20 years?" If we are talking about small/large helicopters, all kinds of helicopters, which would make a difference. So you can get some idea of what kind of system and obviously, this is a much easier question to answer in this kind of situation. If you tried to estimate costs for a particular engine for a particular helicopter, it is easier to define which would mean in terms of the system you are looking at than if you were just trying to look at what kind of generic models would be

available to the community which has to have a general application for anyone that might want to use it.

That kind of gives you the focus of where you intend to go once you end up with a model and decide what it is going to be used for. That is one of the first things that I ask.

IS THIS MORE JUST TO FRAME IT?

It's just to get an idea of what the model is supposed to be used for, because ultimately that's the criteria. Our criteria is always how well does this estimate what we intended to estimate and that is the ultimate criteria; everything else is just needed information to answer that one question. Another thing I would do, and this is something a little bit different, I would look at the cost drivers that they have included or maybe the cost drivers considered and the cost drivers included in the model.

DO THEY USUALLY TELL YOU WHAT FACTORS THEY CONSIDERED OR GIVE YOU WHAT THEY GIVE YOU?

Sometimes somebody will give you their data set and you will have data on variables that they did not end up putting in their model. I think the jet engine model that Rand developed has some variables that they did not include in the model. They came out and said these are the variables we thought might be important, but did not include them.

COULD THERE BE EXTRA SYSTEMS AS WELL?

No. I think in our case they have used all the observations. Obviously the question here is 'is this a complete set or are there some things that you think might have been significant

that we did not consider?' Obviously when you get down here when you look at what was actually brought in, the obvious question is 'why were those variables that they considered to be potential cost drivers not brought in as part of the model? Because they were insignificant; because there were other problems; because of something else? Why were they excluded? Those are other things you would want to think about.

Basically this is just helping you evaluate the extent to which they tackled the identification problem.

The next thing I would want to look at is the model specifications that were considered. In particular, what I look for is something in the documentation that tells me that the model specifications they select were based on some kind of rationale. Sometimes what you'll find is there is no justification of model specification other than best fit to the line. That's the worst possible case and that says they brought no logic into the process at all. The second worst case is that they will develop models that have good statistics and then try to rationalize why those models are acceptable. The best possible case is to come up with your rationale first as to what you think that model might look like and then test those models according to model specification. You'll find them all true.

CAN YOU IDENTIFY FROM THE DOCUMENTATION THAT THEY'VE DONE?

No, it's not always evident. Sometimes they just give you the model, the statistics and that's all. (...) The question of model specification is to look for some kind of rationale as

to whether or not there is some kind of rationale for model specifications and then look at whether it was a rationale look to see if there was some kind of logic to see if it was a quality type of thing or it was just some kind of rationalization that they came up with to justify the models they selected based on some of the criteria on basing the statistics. The best case is to have inquiry logic and the second best case is the logic would have to be a rationale, the worst case would be that you have nothing. That kind of gives me an idea of the thought process that went into putting the structure on it. Once you get past this point then it just becomes a matter evaluating statistics.

DO THEY PUT THIS IN CHAPTERS OR DO THEY JUST (...)?

Normally they don't organize the write up this way. They tend to organize the write up by saying 'here's model A' and they'll talk about model A, then they'll talk about model B, then they'll have the research development costs and then the production costs... Then at the end of that chapter, they'll discuss pretty much everything that they talked about. It really varies a lot. The organization normally is not as critical as just having stuff there. In most cases, you'll read through the report and it'll get 5-6 pages long with maybe an appendix (...) and you'll look for this kind of stuff and won't find it in there anywhere. If they've got it in there I feel pretty good, I don't worry if it's organized or not.

WHEN YOU READ THROUGH, DO YOU HIGHLIGHT, OR JUST LOOK TO SEE

WHAT'S THERE?

I don't have some sort of checklist approach. Typically what I do when I'm reading through it, if I see something that I think is critical in terms of something that I would use to evaluate the model I'll highlight it, check the page or draw a line, to bring me back to that section. Basically all I'm doing there is reading for content and just want to remind myself that when I'm done I want to go back and make sure everything was covered and know that this is where I should go to look for something. This is most of the logic that was up front. Some cost reports just start out with the data or the analysis of the data, which means that none of this is there.

IS THERE A CHANCE OF CALLING THEM AND SAYING 'I'VE GOT A PROPOSAL, I READ IT, I SEE YOUR DATA AND WHAT YOU'VE DONE' IF YOU ASK THEM THESE KINDS OF THINGS DO THEY DO YOU INITIALLY ASK FOR THAT, DOES IT COST YOU MONEY...?

Unfortunately, since a lot of this stuff is judgmental, what you'll have happen if you call up the contractor and say 'did you think about this...' it doesn't matter whether they did or not, they'll say, 'yes, we did' and then they'll give you an answer off the cuff. They're not going to let you know they did a sloppy job. The best way to make sure this gets done is to write it in a statement of work so that they're required to address these issues as part of their analysis.

DO YOU HAVE SOMEONE TO FOLLOW AS FAR A SOW GOES?

No. In fact, my feeling is that if you really wanted to keep the contract honest in doing something like this the other

thing would be to hire them in a two phase process. They can do the (...) logic from microstructure, identification of cost drivers, things like that, and have them submit an additional report before (...) and sit down and evaluate that part and make sure that everybody is in agreement that this is a good structure and criteria to apply and then have them go back and collect and analyze the data. It is just too damn easy once you've got your hands on the data, but I'm not saying you couldn't do this anyway on the side, but once you start to analyze the data to get you thinking to conform to whatever the data tells you, then you can justify it. It doesn't hurt to have them sit down and say 'outline this as a preliminary report, we'll look at it and if we agree with it, go ahead and do your analysis. Obviously, they have to do (...) if not actual data collection, they've got to do something (...)

That doesn't mean down here -- up here, when you're considering what potential (...) it shouldn't make any difference (...) Because one of the things you want to know is if a cost driver that should give an equation is not included because data is not available, you want to know that's happening. That will tell you something about what you can expect that model will do for you. If you've got an influential cost driver that is not included, then you'll end up with non-normal distribution, (...) distributions in their terms.

SO, IF YOU KNOW THAT SOMETHING SHOULD BE INCLUDED AND IT'S NOT, IT'S BECAUSE OF THE AVAILABLE DATA ...



You know that's a weakness in the model and you know that where your (...) but the influence there may be very significant, which could drive your estimate - make your actual cost very different from your estimated cost. Much more so than just random error. That's important enough. (...) AND THEY SAY THEY SUBMIT IT, AND YOU THINK 'HERE'S ONE THEY DIDN'T INCLUDE' AND THEY SAY (...) DATA, THEN THAT'S SOMETHING YOU SAID THEY WOULD SAY 'WELL, WE DIDN'T THINK THAT WAS IMPORTANT'

Well, that's a judgment call because that's logic. If they say they think it's not important and you think it's important, then I guess the question is 'why?'. My feeling is that, and I have found and do the same thing, that once you start getting into the analysis of the data, once you have the data and you are in the situation where you can start analyzing it, you tend to immediately stop thinking about these issues. So, the important thing is to force yourself to do this up front. That's the way to do that. I wouldn't (...) in terms of the contractors to (...) some kind of preliminary study/report that analyzes all of these issues and have it (...) This is the basic logic structure.

BASICALLY YOU'RE JUST GOING TO LOOK, ASSUMING THAT THERE WERE SOME STANDARDS IN THE SOW, THAT THE COST PERSON DIDN'T INFLUENCE OR WENT OUT, OR WHATEVER, THEN WHEN THE MODEL COMES BACK IN YOU LOOK FOR EVIDENCE OF THIS KIND OF THOUGHT PROCESS One of the things - if you're evaluating a product, not evaluating the process - it may turn out that what they did

may be perfectly sound even though they didn't go through all the steps. It might have been pure accident. So, the question is not 'did the contractor do it?' the question is 'if somebody had sat down and done these things, would they have come up with the same answer that the contractor came up with even though he didn't have all the answers to the questions. He may have included a set of cost drivers just because that was the data set that was available. If they turn out, then if you go back, just by sheer luck of the draw, that data set included most of the things that you would have considered to be significant cost drivers. In that case, the contractor is O.K., even though we never even bothered to ask the question; even though you were supposed to, he still turned out O.K. If you're analyzing a cost model, the question to ask is not 'did the contractor do it?'; the question to ask is 'if somebody went through this process, would they get to the same point that the contractor got to?' You may be asked to evaluate these answers by going to some source (...) Once you get to this point, then basically you can start getting into a data analysis (...) I guess the next thing I would want to get, and sometimes this is a little bit harder to tell, is looking at the data itself in terms of quality (for lack of a better word). Are there any problems that you can tell in terms of the data that they used? This is an extremely difficult question to answer and then back up. You could never really know for sure that the data is correct. You can have a source, find out a history of the programs in

that data set, just to get an idea of whether there is anything that happened in the programs that might have been different from what normally occurs in a normal program. Things like (...) among other things look at would be a problem with that jet engine, and some of those engines are new developments and some are derivatives. If you're looking at research development costs, does that make a difference? Does the development cost of the B2-B really represent (...) You might want to look at some things like that. Of course, they could go the other way too. You might have a development like the ATF and say that the ATF is really extending the state of the art well beyond the existing level of technology, much more so than the fighter aircraft has ever done. That's probably not true, but not every fighter aircraft necessarily extend the same error to the same degree. The F-111 was probably an aircraft that really (...) the state of the art (...) The F-104 probably didn't push it as hard. The problem with answering those kinds of questions is that you have to know something about the programs, the events that lie behind that data, and that's not always easy to find out. (...) If you can go back to somebody in the program office and ask what really went on in that program. The F-16 had the multi-national production, that kind of thing; does that make that cost typical or atypical for the aircraft? That's why it's so difficult to answer that question on that kind of issue. It's a history lesson. Another difficult question to ask about is making sure that the data is consistent.

Normally when you're helping a cost officer comparing cost (...) accounting systems for different contractors, contractors do not have any kind of stand in the way of accounting for costs. In fact, one contractor probably doesn't have a standard way of accounting for costs (...) all the time. You can look at an aircraft built by McDonel Douglas in the 60's and (...) in the 80's and their accounting system can be totally different. So, you never really (...) those effects, but you want to make a small amount that they (...) Those are really tough questions to answer.

Usually what happens here is that people go ahead and analyze the data and if they get reasonably good fits, they assume that's not a problem. They are almost saying that a problem with the data is going to show up because of points (...) or something like that. If everything seems to be "well behaved" then you assume that the data is reasonably good. The only problem is with small data sets, sometimes you can fit equations that seem to be well behaved even though at some point they can be highly distorted.

WHAT IS THE SIZE OF THE AVERAGE DATA SET(...)?

Most of my data sets deal with major record systems. I would say I work with data sets for 10-25...

IS THERE A MINIMUM THAT YOU WOULD TURN A MODEL BACK FOR?

Suppose that's all that's there. Suppose that there isn't anything else. It still (...) Obviously you wouldn't put a lot of confidence in it, and you'd take that into account. But still, if that's all the information you've got, that's

still better than not knowing anything. If you don't accept that data then you end up (...) If it was really small or really erratic, you might have to say that. (...) what I need is an expert. I've seen one model that was developed on three data points and (...)

WAS IT SOMETHING THAT WAS SUBMITTED TO US...?

No. This was a CER that was developed by a contractor in-house. The reason that he had such a small data set was that because of the data points he had (...) the system (...) to develop. (...) data points with 10 or 15 systems you look at it like jet engines; half are (...), half are GE and that's not even true (...), so if I was a GE (...) that model, I'd have it less than half the days that you had it. Especially when contractors try to up their own cost estimate relationship (...)

WHAT CONTRACT (...)

(...)DOES RAND STILL USE THIS?

Rand doesn't do it very much anymore. They consider this too pedestrian for them. They like to look at costs (...) where you're having to define the system. Once it gets to the point where someone says (...) At this point, not even knowing anything about the model, one of the things you can do is that since it doesn't involve the model you can look at an outline of this (...) before you know the model (...) Very few people do that.

IS IF USEFUL?

I think it is. The reason I say that is because sometimes you

may find yourself rejected. The problem, is that when you get into analysis of the data, is sometimes you reject models because the statistics are so terrible and you don't bother to go on and look at all the diagnostics. It may be that the only reason that the statistics are terrible is because of the influence of a couple of outliers. In fact they were legitimate and you dealt with them and that model may be a fantastic model. But, you may end up not even considering it because you don't get past the first stage. Why spend time looking at a model that on the surface appears to be a terrible (...) It's probably worthwhile looking at this, and the key thing you're looking at here at this point is (...) At this point (...) the only question you can ask is 'is this a legitimate (...) You're looking for gaps. You may have values of (...) within your relevant range where which really don't have enough data to really see what's happening to the model. (...) and the question is what's happening here; what's going on in here between these two groups of data. If you don't have any information it's almost like extended beyond the (...) and (...) But, still, you're (...) somewhere between here and here (...) well behaved venture. (...)THESE AREN'T NECESSARILY POINTS TO REJECT IT, IT'S JUST INDICATING...

It's just indicating that in this particular model (...) there are going to be some weaknesses since I don't know what's going on, especially if I assume continuity between data points. You don't see that in reference analysis. You'll see

people identifying the (...) points of the analysis (...) if you go beyond this range you may be in trouble because you're using some assumptions which may be invalid. (...)

Once you get past this then you're ready to start (...)

IS THERE ANYTHING ELSE THAT YOU WOULD DO BEFORE YOU SIT DOWN TO (...)

(...) There is a tendency on the part of people that are statistically oriented to want to crunch numbers and yet there's 29 other programs out there that do the same thing.

IF I WERE TO GET A MODEL AND GO TO A COMPUTER, WHAT WOULD BE THE MOST USEFUL METHOD FOR THIS INFORMATION TO COME AT ME - SHOULD IT BE COMPUTER PROMPTED? WE'LL JUST MAKE YOU THE COMPUTER - HOW WOULD YOU PROMPT ME TO EVALUATE IT?

I guess that at the initial point probably the first thing I would do is to (...) questions. I'll be right up front with you, I don't know very much about expert systems in terms of (...) It's easier (...) computer (...) no responses (...)

JUST KNOW WE CAN DO A MENU OF PARAGRAPHS FOR WHICH ONE FITS THE SITUATION, WE CAN DO HYPERTEXT, LIKE 'IF IT'S THIS, WE'LL TAKE YOU HERE, IF IT'S THIS, WE'LL TAKE YOU HERE, THAT KIND OF THING. WE CAN DO CALCULATIONS INTERNALLY - GETTING A RANDOM VALUE FOR THIS IF THE VALUE IS BETWEEN 'THIS' AND 'THIS'.

The first part of it question one is basically

WOULD THAT INFORMATION IF THAT WERE SAVED, WOULD I USE THAT LATER SOMETIME, OR IS IT JUST TO FRAME...?

Only when you get down to cost drivers and suppose you say, technology is an important cost driver, and that no variable

captures the influence of technology, the question is (...) look at the technology that reflecting your data base and the technology is going to be incorporated into your system. Ask if there is significant change in technology, because even though technology is a significant cost driver, it isn't any different (...) The question doesn't always have an impact on cost, but what you have to ask is this a value that's changing over observations (...) or is it going to be very different for (...) Now those can be the answers to the yes/no questions. If the answer is that the technology isn't really changing that much between the data and the observation (...) in my new system, then you would know this was a different cost driver and you don't need to include it. You can incorporate that into your definition of (...)

ANOTHER THING EXPERT SYSTEMS DO NOW - IT CAN HANDLE UNKNOWNNS IS THIS LIKE A QUANTUM JUMP IN TECHNOLOGY? IF YOU PUT "UNKNOWN", "?", ETC., IT CAN HANDLE THAT. WHAT IT DOES IS ASSIGNS CONFIDENCE FACTORS IT WILL ASSIGN EACH RESPONSE SO IF YOU PUT "UNKNOWN" THEN YOU CAN ASSIGN A CERTAIN CONFIDENCE FACTOR TECHNOLOGY THAT HAS VERY LITTLE CONFIDENCE FACTOR YOU CAN PLAY WITH THAT. YOU CAN HANDLE A CERTAIN AMOUNT OF AMBIGUITY WITH RESPONSES.



## Appendix B: Program Code

NOTE: The following appendix contains the program code for the Cost Model Evaluation Program. Files start at the top of a page with an underlined file name and extension. Any explanation of the file purpose is noted in italics.

### START.BAT

This batch file runs the different VP-Expert files in order. The batch file is executed by typing "Start" at the computer prompt once all the files and programs have been loaded. "VPX" is the execution command for VP-Expert.

VPX INTRO  
VPX HYPER  
VPX TYPELIST  
VPX SIGLEVEL  
VPX EQUATION  
VPX SETSIZE  
VPX OVERLOOK  
VPX SOURRAW  
VPX RELEVANT  
VPX HOMOGENE  
VPX OUTWRTX  
VPX VARINFO  
VPX LISTCON  
VPX OVERCON  
VPX RELEVCON  
VPX FACTRCON  
VPX IDENCON  
VPX SPECCON1  
VPX SPECCON2  
VPX ANALYSIS

## FILES

This ASCII file lists the names of the files down the left side and the file extension across the top. The column labeled "KBS" are VP-Expert files. The column labeled "FACTS" are ASCII files that contain the saved knowledge from the associated VP-Expert file. The prefix "L-" indicates the facts are loaded while the prefix "S-" indicates the facts are saved to that filename. The column labeled "TXT" is the extension for a hypertext information file. The column labeled "DBF" indicates database files and the column labeled "BAT" indicates batch files.

KBS	FACTS	TXT	DBF	BAT
INTRO				
HYPER		YAKBASE		
TYPELIST	S-TYPELIST		AYAKBASE YAKBASE EMPTY DEFINE	COPYONE
SIGLEVEL	S-SIGLEVEL			
EQUATION	L-TYPELIST S-EQUATION		DEFINE EMPTY EQUATION	COPYTWO
SETSIZE	S-SETSIZE			
OVERLOOK	S-OVERLOOK			
SOURRAW	S-SOURCE S-RAWDATA		EQUATION	
RELEVANT	L-SETSIZE S-RELEVANT		EQUATION	
HOMOGENE	L-EQUATION S-HOMOGENE		EQUATION	
OUTWRTX	L-EQUATION S-OUTWRTX		OUTWRTX	COPYTHREE
VARINFO			YAKBASE EQUATION DERIVE	

KBS	FACTS	TXT	DBF	BAT
LISTCON	L-TYPELIST S-LISTCON			
RELEVCON	L-RELEVANT L-EQUATION S-RELEVCON			
OVERCON	L-OVERLOOK			
FACTRCON	L-EQUATION		AYAKBASE	
IDENCON			YAKBASE EQUATION	
SPECCON1	L-EQUATION L-SETSIZE			
SPECCON2			YAKBASE EQUATION	
ANALYSIS	L-SIGLEVEL			

INTRO.KBS

```
!-----INITIAL SET UP-----
    ENDOFF;
    EXECUTE;
    RUNTIME;
    BKCOLOR = 3;
!-----ACTION BLOCK-----
ACTIONS
DISPLAY ""
DISPLAY ""
DISPLAY ""
COLOR = 4
DISPLAY "          WELCOME TO THE COST MODEL EVALUATION
EXPERT SYSTEM"
COLOR = 0
DISPLAY ""
DISPLAY "This system is set up to help you evaluate a cost
model that"
DISPLAY "          has been submitted by an outside
source."
DISPLAY ""
DISPLAY ""
DISPLAY "          Sit down and relax - The program will take
about 30 minutes."
DISPLAY ""
COLOR = 0
DISPLAY "          Observe the bottom of the screen for input
instructions"
DISPLAY "          written in GREEN LETTERS."
DISPLAY ""
DISPLAY ""
DISPLAY ""
FIND CONTINUE
RESET CONTINUE
CLS;

ASK CONTINUE:
" TO CONTINUE,";
CHOICES CONTINUE:
___PRESS_RETURN___;
```

## HYPER.KBS

```
!-----INITIAL SET UP-----
    ENDOFF;
    EXECUTE;
    RUNTIME;
    BKCOLOR = 3;
!ACTIONS
BLOCK-----BACKGROUND-----HYPERTEXT
INFORMATION
ACTIONS
CLS
COLOR = 4
DISPLAY"                                BACKGROUND INFORMATION"
DISPLAY""
COLOR = 0
DISPLAY"        Assumptions and background information are
contained in this"
DISPLAY"                                first section."
DISPLAY""
COLOR = 20
DISPLAY"        YOU WILL NEED A MOUSE DEVICE TO PROPERLY USE
THIS SECTION"
DISPLAY""
COLOR = 15
DISPLAY"If a mouse is not hooked up to your computer, you can
still run the"
DISPLAY" rest of the program.  Simply strike the ESC key when
the black"
DISPLAY"                                background screen appears."
COLOR = 0
DISPLAY""
DISPLAY""
DISPLAY"The information in this section IS NOT PRINTED OUT.
You may want"
DISPLAY"to write down any information / definitions that you
feel you need"
DISPLAY"after this section is completed. Another option is hit
the"
DISPLAY"PRINT SCREEN when you want to print the screen
display."
DISPLAY""
FIND CONTINUE
RESET CONTINUE

GMODE 6
GBCOLOR 3
MOUSEON
EXIT = NO
FIND LEVEL
WHILETRUE EXIT = NO THEN
END;
```

!RULES BLOCK-----BACKGROUND-----Hypertext Rules

RULE 1

IF EXIT = NO

THEN LEVEL = YES;

WHENEVER 1

IF LEVEL = YES

THEN HYPER1= INSTRUCTIONS

RESET INSTRUCTIONS;

!STATEMENTS

BLOCK---BACKGROUND-----

ASK CONTINUE:

" TO CONTINUE,";

CHOICES CONTINUE:

\_\_\_\_PRESS\_RETURN\_\_\_\_;

HYPertext HYPER1: 1,1,76,22,YAKBASE,7,0;

LBUTTON EXIT: 65,24,7,0,EXIT;

## YAKBASE.TXT

### \*INSTRUCTIONS

This is the explanatory portion of this program.

Exit -----To exit at any time, position the mouse  
over the "EXIT" button and click.

Continue -----To continue, follow the text on the screen.

Hypertext Words -Capitalized words are hypertext  
words (for example - Purpose ).

-Positioning the mouse over these words and  
Clicking will open a window providing more  
information.

-Clicking on the word INDEX will return you  
to the main index.

Start by clicking here -> Index

### \*INDEX

Hyperscreen - Master Index

This index will allow you to move to the topics you want to  
review.

- |                                      |                 |
|--------------------------------------|-----------------|
| 1. -> Purpose                        | 4. Definitions  |
| 2. -> Assumptions                    | -> sow          |
| 3. -> Proper_Sow                     | -> rfp          |
| Factors_that_influence_cost          |                 |
| Key_cost_drivers                     | 5. Instructions |
| Identification_of_cost_drivers       |                 |
| Specification_logic                  |                 |
| Raw_data                             |                 |
| Model_Properties_and_Characteristics |                 |
| Outliers                             |                 |
| Ommitted_Variables                   |                 |
| Heteroscedasticity                   |                 |
| Normality_Of_Residuals               |                 |
| Autocorrelation                      |                 |

Click Here to continue -> Purpose

### \*PURPOSE

Hyperscreen - Purpose of a Cost Model

This system will help you explore the validity of a cost  
model.

The validity of a cost model is based on the following question:

"How Well Does The Model Predict What We Want To Predict ?"

This is the ultimate question that the analyst will have to answer

when evaluating the usefulness of a cost model.

[Click Here to continue](#) -> Assumptions

[Click Here for the index](#) -> Index

#### \*ASSUMPTIONS

Hyperscreen - Assumptions of this System

Assumptions for the use of this model are listed below

- 1 - A properly completed Statement of Work or Sow
- 2 - A Request for Proposal or Rfp that forces adherence to the Sow
- 3 - Availability of all data and statistics

The first of these assumptions assures the second and third.  
The contents of a properly completed Sow follow:

[Click Here to continue](#) -> Proper\_Sow

[Click Here for the index](#) -> Index

#### \*PROPER\_SOW

Hyperscreen - Requirements for a Proper\_Sow completion

A properly completed Sow should request the following information:

1. A list of all the Factors\_that\_influence\_cost
2. A list of all the Key\_cost\_drivers
3. A discussion of Identification\_of\_cost\_drivers
4. A discussion of Specification\_logic
5. A list of all the Raw\_data
6. A discussion of Model\_Properties\_and\_Characteristics
7. A discussion of Outliers
8. A discussion of Ommitted\_Variables
9. A discussion of Heteroscedasticity
10. A discussion of Normality\_Of\_Residuals
11. A discussion of Autocorrelation

[Click Here to continue](#) -> Factors\_that\_influence\_cost

[Click Here for the index](#) -> Index

#### \*FACTORS\_THAT\_INFLUENCE\_COST



Hyperscreen - A discussion of factors\_that\_influence\_cost

A list should be included of all factors that could influence the cost of the population. Each factor identified should be captured by a cost driver in the model.

Click Here to continue -> Key\_cost\_drivers

Click Here for the index -> Index

#### \*KEY\_COST\_DRIVERS

Hyperscreen - A discussion of key\_cost\_drivers

The key cost drivers are variables which have a specific behavior with respect to cost. These cost drivers are said to "capture" the change in cost.

Click Here to continue -> Identification\_of\_cost\_drivers

Click Here for the index -> Index

#### \*IDENTIFICATION\_OF\_COST\_DRIVERS

Hyperscreen - Identification\_of\_cost\_drivers

From a list key cost drivers, an equation is built using the least squared best fit method. This method may indicated certain key cost drivers better explain the variation in cost than other key cost drivers. The cost analyst must then assess which variables to include and exclude from the equation.

This item includes           a. Why cost drivers were excluded  
                                  b. Why cost drivers were included

Click Here to continue -> Specification\_logic

Click Here for the index -> Index

#### \*SPECIFICATION\_LOGIC

Hyperscreen - An explanation of specification\_logic

The specification logic deals with the exact form the variable assumes in the cost model equation. Variables may be transformed to indicate different cost behavior. Hence, a premeditated logic should be included explaining the form of a cost driver in the equation.

This includes           a. The ranges expected  
                                  b. An explanation of cost behavior

Click Here to continue -> Raw\_data  
Click Here for the index -> Index

#### \*RAW\_DATA

Hyperscreen - Issues of raw\_data

The data used to create the cost model should be included with the cost model itself. The raw data must be listed with background information.

This includes

- a. An explanation of any adjustments  
(for example: inflation procedures)
- b. Adjusted data listings
- c. Identification of programs for each data point
  - 1. Including a brief history
  - 2. Any events that may have impacted  
cost of the program  
(for example: A Labor Strike)
  - 3. Description of the accounting system

Click Here to continue ->  
Model\_Properties\_and\_Characteristics  
Click Here for the index -> Index

#### \*MODEL\_PROPERTIES\_AND\_CHARACTERISTICS

Hyperscreen - Elements of  
Model\_properties\_and\_characteristics

This includes

- a. Model behavior over the range of the data
- b. Significance levels expected for
  - 1. F-Tests
  - 2. T-Tests
  - 3. R-Square
  - 4. Coefficient of Variation

Click Here to continue -> Outliers  
Click Here for the index -> Index

#### \*OUTLIERS

Hyperscreen - A discussion on outliers

A discussion of outliers with respect to the independent and the dependent variables must be included. Several methods are available to quantify a data point as an outlier. These will be covered in the cost model.

Click Here to continue -> Ommitted\_Variables  
Click Here for the index -> Index

## **\*OMMITTED\_VARIABLES**

Hyperscreen - A discussion on Ommitted\_variables

If a variable is ommitted, the cost model may not be capable of capturing some cost. This increases the potential of a bad estimate. Some possible reasons for omitting variables should be considered.

These include

- a. Considerations of lack of data
- b. Collinearity discussions
  - 1. Among model variables
  - 2. Among ommitted variables
- c. Statistical insignificance discussions

Click Here to continue -> Heteroscedasticity  
Click Here for the index -> Index

## **\*HETEROSCEDASTICITY**

Hyperscreen - A discussion of Heteroscedasticity

The condition of the error variance not being constant over all cases is called heteroscedasticity, in contrast to the condition of equal error variance, called homoscedasticity.

(Reference: Applied Linear Regression Models by John Neter - page 423)

Click Here to continue -> Normality\_Of\_Residuals  
Click Here for the index -> Index

## **\*NORMALITY\_OF\_RESIDUALS**

Hyperscreen - A discussion of Normality\_of\_Residuals

Normality plots of the residuals are plots which put each residual against its expected value when the distribution is normal. A plot that is nearly linear suggests agreement with normality, whereas a plot that departs substantially from linearity suggests that the error distribution is not normal.

(Reference: Applied Linear Regression Models by John Neter - page 125)

Click Here to continue -> Autocorrelation  
Click Here for the index -> Index

## **\*AUTOCORRELATION**

## Hyperscreen - A discussion of Autocorrelation

One of the assumptions of basic regression models is that the random error terms are either uncorrelated random variables or independent normal random variables. In some applications, regression involves time series data. For such data, the assumption of uncorrelated or independent error terms is often not appropriate; rather, the error terms are frequently correlated positively over time.

Error terms correlated over time are said to be autocorrelated or serially correlated.

(REFERENCE: Applied Linear Regression Models by John Neter - page 484)

[Click Here for the index -> Index](#)

\*SOW

## Hyperscreen - A discussion of the SOW

The Statement Of Work or SOW is a section of a Request for Proposal.

This section specifies what tasks are required for the proper completion of a contract. The contractor is expected to price these

tasks and respond to the Request for Proposal with a package explaining his method of accomplishing the items set out in the SOW.

[Click Here for the index -> Index](#)

\*RFP

## Hyperscreen - A discussion of the RFP

The Request for Proposal is a way the government can solicit priced bids for work. These bids can then be evaluated and a choice can be made based on the most cost effective option.

[Click Here for the index -> Index](#)

TYPELIST.KBS

!ACTIONS BLOCK---COST DRIVERS CONSIDERED-----  
ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS  
COLOR = 1

```

        LOOP = 1
        SWITCH_Definition = EDIT
    WHILETRUE SWITCH_Definition = EDIT
        THEN BCALL COPYONE
            RESET          Listed_Factors
            RESET          More_Listed_Factors
            RESET TRIGGER_More_Listed_Factors
            RESET    LIST_More_Listed_Factors
        COLOR = 0
    DISPLAY"-----"
    COLOR = 4
    DISPLAY "                                CONSIDERATION OF COST DRIVERS"
        COLOR = 20
        DISPLAY "                                Iteration #"
    {LOOP}"
        COLOR = 0
        LOOP =(LOOP + 1)
        MENU System_Type,ALL,YAKBASE,Type
        FIND System_Type
        DISPLAY""
    COLOR = 0
    DISPLAY
    "-----"
    -----"
    COLOR = 4
    DISPLAY "    COST DRIVERS FROM THE EXPERT'S LIST AND THE
    CONTRACTOR'S LIST"
    COLOR = 0
    DISPLAY ""
        MENU
    Listed_Factors,System_Type=Type,YAKBASE,Driver
        FIND          Listed_Factors
        COUNT
    Listed_Factors,COUNT_Listed_Factors
        FIND CONTINUE
        RESET CONTINUE
        CLS

    COLOR = 0
    DISPLAY"-----"
    -----"
    COLOR = 4
```

```

DISPLAY"                COST DRIVERS FROM THE CONTRACTOR'S LIST
ONLY"
COLOR = 0
DISPLAY"-----"
-----"
                FIND                More_Listed_Factors

CLS
COLOR = 0
DISPLAY"-----"
-----"
COLOR = 4
FIND CONTINUE
RESET CONTINUE
COLOR = 1
RESET SWITCH_Definition
FIND SWITCH_Definition
END
CLS

DISPLAY"Please wait while the program updates the database."

SAVEFACTS TYPELIST

WHILEKNOWN    POPPER_Listed_Factors
                RESET    POPPER_Listed_Factors
                POP        Listed_Factors, POPPER_Listed_Factors
                GET        POPPER_Listed_Factors = Driver,YAKBASE,ALL
                        In_List = YES
                PUT    YAKBASE
                CLOSE YAKBASE

                END

WHILEKNOWN    RECORD_NUM
                RESET RECORD_NUM
                SET_In_List = YES
                GET SET_In_List = In_List AND
                System_type = Type,YAKBASE, ALL
                APPEND DEFINE
                CLOSE DEFINE
END;

!RULES BLOCK-----TYPE AND
LIST-----

RULE CONTRACTOR_ONLY
IF TRIGGER_More_Listed_Factors = YES
THEN    More_Listed_Factors = YES
        FIND        LIST_More_Listed_Factors
        WHILEKNOWN    LIST_More_Listed_Factors
                FIND        Include_Factor
                RESET    LIST_More_Listed_Factors
                RESET    TRIGGER_Include_Factor
                RESET    Include_Factor

```

```

FIND      LIST_More_Listed_Factors
END
ELSE      More_Listed_Factors = NO
DISPLAY "No cost drivers are listed by the contractor only.";

```

```

RULE TRIGGER_INCLUDE_FACTOR
IF TRIGGER_Include_Factor = YES
THEN      Include_Factor = YES
          Driver   = (LIST_More_Listed_Factors)
          Type     = (System_Type)
          In_List  = YES
          New_Var  = YES
          Composite= ?
          Stepindc = ?
          Sign     = ?
          Derivel  = ?
          Derive2  = ?

```

```

      APPEND YAKBASE
ELSE
      Include_Factor = NO
          Driver   = (LIST_More_Listed_Factors)
          Type     = (System_Type)
          In_List  = NO
          New_Var  = YES
          Composite= ?
          Stepindc = ?
          Sign     = ?
          Derivel  = ?
          Derive2  = ?

```

```

      APPEND YAKBASE;

```

```

!STATEMENTS BLOCK-----PART 1 - POPULATION /DEFINITION---ASK
<>CHOICES <> PLURAL

```

```

      ASK CONTINUE:"          TO CONTINUE";
      CHOICES CONTINUE:____PRESS_RETURN____;

```

```

      ASK      System_Type:
"      What is the general nature of the system in question?";

```

```

"      ASK      Listed_Factors :

```

```

Pick the cost drivers from the list below that are contained
      in the contractor's list of cost drivers.

```

```

";

```

```

      ASK      TRIGGER_More_Listed_Factors :
"Are there any other cost drivers in the contractors list?

```

```

";

```

```

      CHOICES TRIGGER_More_Listed_Factors :YES,NO;

```

```

      ASK      LIST_More_Listed_Factors :
"List other cost drivers that are included in the contractors

```

list.  
Enter a ? as your last list entry";

ASK TRIGGER\_Include\_Factor:  
"Should I include this cost driver for consideration  
even though the experts do not recognize it as one?";  
CHOICES TRIGGER\_Include\_Factor:YES,NO;

ASK SWITCH\_Definition:  
"Choose ? if you are done or EDIT to change your inputs.";

PLURAL: Type,  
Listed\_Factors;



COPYONE.BAT

@ ECHO OFF

REM THIS BATCH FILE MAKES WORKING FILES FROM "A" (A =  
ARCHIVE) FILES

COPY AYAKBASE.DBF YAKBASE.DBF

REM CREATES THE WORKING DBF FROM THE MASTER DATABASE

COPY EMPTY.DBF DEFINE.DBF

REM CREATES THE WORKING DATABASE FOR THE DEFINITION PORTION

SIGLEVEL.KBS

!ACTIONS BLOCK-----LEVEL OF ACCEPTANCE FOR MODEL  
STATISTICS-----

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;  
ACTIONS

CLS  
COLOR = 4  
DISPLAY "-----"  
-----"  
COLOR = 0  
DISPLAY "           LEVEL OF ACCEPTANCE FOR MODEL STATISTICS"  
COLOR = 15  
DISPLAY " You must decide the acceptance levels for model  
statistics"  
DISPLAY "   you will use to determine if the equation is  
significant."  
DISPLAY "The level of acceptance is related to the  
significance level."  
DISPLAY "   The significance level is the Type I error  
probability."  
COLOR = 4  
DISPLAY "           LEVEL OF ACCEPTANCE = 1 - Type I error  
probability"  
COLOR = 0  
DISPLAY "   Some equations can be accepted with lower  
statistics if"  
DISPLAY "   the included variables are deemed crucial to the  
model."  
DISPLAY "   You should record two different levels of  
acceptance."  
DISPLAY ""  
DISPLAY ""  
COLOR = 4  
FIND CONTINUE  
RESET CONTINUE  
CLS  
COLOR = 0  
FIND ACCEPTANCE\_Normal  
FIND ACCEPTANCE\_Exceptions  
DISPLAY ""  
COLOR = 4  
FIND CONTINUE  
RESET CONTINUE

SAVEFACTS SIGLEVEL;

!RULES BLOCK-----LEVEL OF ACCEPTANCE FOR MODEL STATISTICS

!STATEMENTS BLOCK---LEVEL OF ACCEPTANCE---

```
      ASK      ACCEPTANCE_Normal:
"What is the level of acceptance for model statistics
(percentages)
  for an equation when all variables are considered equally?
    (Your answer should be between 50 and 100)";
      RANGE    ACCEPTANCE_Normal: 50,100;

      ASK      ACCEPTANCE_Exceptions:
"What is the level of acceptance for model statistics
(percentages)
  for an equation when certain variables must be included?
    (Your answer should be between 50 and 100)";
      RANGE    ACCEPTANCE_Exceptions: 50,100;

      ASK CONTINUE:"      TO CONTINUE";
CHOICES CONTINUE:___PRESS_RETURN___;
```

## EQUATION.KBS

!ACTION BLOCK-----EQUATION VARIABLE

INPUT-----

ENDOFF;

EXECUTE;

RUNTIME;

BKCOLOR = 3;

ACTIONS

LOADFACTS TYPELIST

LOOP\_Input = 1

SWITCH\_Variable\_Input = EDIT

WHILETRUE SWITCH\_Variable\_Input = EDIT

THEN BCALL COPYTWO

RESET EQUATION\_Variable\_Input

COLOR = 0

DISPLAY

"-----"

-----"

COLOR = 4

DISPLAY "

EQUATION VARIABLE SELECTION"

DISPLAY "

Iteration # {LOOP\_Input}"

COLOR = 0

DISPLAY

"-----"

-----"

COLOR = 15

DISPLAY"THIS MODEL ASSUMES EACH VARIABLE APPEARS IN THE  
EQUATION ONLY ONCE."

LOOP\_Input = (LOOP\_Input + 1)

MENU EQUATION\_Variable\_Input, System\_Type =

(Type), DEFINE, Driver

FIND EQUATION\_Variable\_Input

COUNT

EQUATION\_Variable\_Input, COUNT\_Equation\_Variables

RESET SWITCH\_Variable\_Input

FIND SWITCH\_Variable\_Input

CLS

END

CLOSE DEFINE

COLOR = 4

DISPLAY "Please wait while the program updates the database"

SAVEFACTS EQUATION

WHILEKNOWN POPPER\_Equation\_Variable

RESET POPPER\_Equation\_Variable

POP EQUATION\_Variable\_Input,

POPPER\_Equation\_Variable

GET POPPER\_Equation\_Variable = Driver, YAKBASE, ALL

```

        In_Eq = YES
        PUT YAKBASE
        CLOSE YAKBASE
    END

    WHILEKNOWN RECORD_NUM
        RESET RECORD_NUM
        SET_In_Eq = YES
        GET SET_In_Eq = In_Eq, YAKBASE,ALL
        APPEND EQUATION
        CLOSE EQUATION
    END;

    !-----RULES BLOCK-----EQUATION VARIABLE
    INPUT-----

    !STATEMENT BLOCK-----EQUATION VARIABLE
    INPUT-----

    "        ASK      EQUATION_Variable_Input:

    "
    Pick the variables from this list that are contained in the
    equation.
    ";

    "        ASK      SWITCH_Variable_Input:
    "Choose ? if you are done or EDIT to change the previous
    variable inputs.";

    PLURAL:      EQUATION_Variable_Input;
    AUTOQUERY;

```

COPYTWO.BAT

@ ECHO OFF  
COPY EMPTY.DBF EQUATION.DBF

## SETSIZE.KBS

!ACTIONS BLOCK----DATA ANALYSIS-----SET SIZE

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS  
COLOR = 0

CLS

DISPLAY"-----  
-----"

COLOR = 4

DISPLAY " DATA ANALYSIS "

COLOR = 0

DISPLAY

"-----  
"

COLOR = 1

DISPLAY "This section of the program requires you to have all  
the data"

DISPLAY "possible on each data point. "

DISPLAY ""

FIND CONTINUE

RESET CONTINUE

CLS

COLOR = 0

DISPLAY

"-----  
-----"

COLOR = 4

DISPLAY "DATA SET SIZE"

COLOR = 0

DISPLAY

"-----  
-----"

FIND DATA\_Number\_of\_Points

DISPLAY ""

DISPLAY "NUMBER OF DATA POINTS = {DATA\_Number\_of\_Points}"

DISPLAY ""

FIND CONTINUE

RESET CONTINUE

SAVEFACTS SETSIZE;

!RULES BLOCK-----SET

SIZE-----

!STATEMENTS BLOCK-----SET

SIZE-----

```
      ASK      DATA_Number_of_Points:
" How many data points are there in the data set provided?";
      RANGE    DATA_Number_of_Points: 1,10000;

ASK CONTINUE:      "
  TO CONTINUE";
CHOICES CONTINUE: __PRESS_RETURN__;
```



# OVERLOOK.KBS

!ACTIONS BLOCK-----OVERLOOKED DATA

POINTS-----

ENDOFF;

EXECUTE;

RUNTIME;

BKCOLOR = 3;

ACTIONS

COLOR = 15

  DISPLAY ""

  DISPLAY "A QUESTION ABOUT OVERLOOKED MEMBERS OF THE DEFINED  
PCPULATION"

  DISPLAY ""

COLOR = 0

  FIND TRIGGER\_Overlook

  DISPLAY ""

  SAVEFACTS OVERLOOK;

!STATEMENTS

BLOCK-----OVERLOOKS-----

--

  ASK       TRIGGER\_Overlook:

"Are there other systems or data points that  
could have been included in the data set?"

  ";

  CHOICES TRIGGER\_Overlook: YES, NO;

SOURRAW.KBS

!ACTIONS

BLOCK-----SOURRAW-----

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS

COLOR = 0

DISPLAY

"-----  
-----"

COLOR = 4

DISPLAY "DATA INTEGRITY"

COLOR = 0

DISPLAY

"-----  
-----"

MENU Good\_Data\_Source,ALL,EQUATION,Driver

DISPLAY"For each variable in the equation, there is the  
question of"

DISPLAY" data integrity. For example,"

COLOR = 1

DISPLAY" Did the accounting systems provide cost information  
for these"

DISPLAY"data points from similar systems using acceptable  
accounting methods?"

COLOR = 0

DISPLAY" OR"

COLOR = 4

DISPLAY" Did the accounting systems provide cost information  
for these"

DISPLAY" data points were obtained from different systems or  
by use of"

DISPLAY" unacceptable accounting principles."

COLOR = 15

DISPLAY"This is a difficult question to answer, but look at  
what information"

DISPLAY" you can and attempt to make a determination of  
confidence in the"

DISPLAY" data for each variable."

COLOR = 0

DISPLAY" Look at the data sources for each variable."

FIND Good\_Data\_Source

COUNT Good\_Data\_Source, COUNT\_Good\_Data\_Source

LOOP\_SOURCE = 1

WHILETRUE COUNT\_Good\_Data\_Source >= (LOOP\_SOURCE)

```

THEN
LOOP_SOURCE = (LOOP_SOURCE + 1)
  POP      Good_Data_Source, POPPER_Good_Data_Source
    GET      Driver =
(POPPER_Good_Data_Source), EQUATION, ALL
    Source = GOOD
  PUT      EQUATION
END

```

```

FIND CONTINUE
RESET CONTINUE
SAVEFACTS SOURCE
CLS

```

```

COLOR = 0
DISPLAY

```

```

"-----"
-----"

```

```

COLOR = 4
DISPLAY "RAW DATA AND DATA ADJUSTMENTS"
COLOR = 0
DISPLAY

```

```

"-----"
-----"

```

```

DISPLAY""
DISPLAY"For most models, the raw data should be required. Of
course, if "
DISPLAY"you did not get the raw data, you will not be able to
validate"
DISPLAY"any changes made to it."
DISPLAY""
FIND DATA_Adjustments
DISPLAY ""
FIND CONTINUE
RESET CONTINUE
CLS

```

```

SAVEFACTS RAWDATA;

```

```

!RULES
BLOCK-----
-----

```

```

      RULE      DATA_Adjustments
      IF      TRIGGER_DATA_Adjustments = YES
      THEN      DATA_Adjustments = YES
                COLOR = 1

```

```

DISPLAY "When data is adjusted, the method must be obvious and
acceptable."
DISPLAY "If you did not get the data, you can make no
assumptions here."
DISPLAY "Inflation indices are a common data adjustment.  In
this case,"

```

```

DISPLAY "indices must be provided and applied using an
acceptable procedure."
DISPLAY "Another common adjustment is for differences in
quantity."
                                DISPLAY ""
                                COLOR = 4
DISPLAY "                                IF THIS IS NOT THE CASE, THE PROBABILITY
FOR"
DISPLAY "                                ESTIMATING ERROR MAY BE GREATER THAN
NORMAL."
                                COLOR = 0
                                ELSE DATA_Adjustments = NO
                                COLOR = 1
DISPLAY "The Raw Data was included. No adjustments were
made."
                                COLOR = 0
                                DISPLAY "";

```

!STATEMENTS

BLOCK-----  
-----

```

                                ASK Good_Data_Source:
"Pick the variables below that you feel come from acceptable
data sources.
";

                                ASK TRIGGER_DATA_Adjustments:
"
Was the raw data adjusted in any way? (If you don't have data,
answer YES)";
                                CHOICES TRIGGER_DATA_Adjustments: YES,NO;

ASK CONTINUE: "
TO CONTINUE";
CHOICES CONTINUE: __PRESS_RETURN__;

PLURAL:Good_Data_Source;

```

## RELEVANT.KBS

```
!ACTIONS BLOCK-----RELEVANT
RANGE-----
ENDOFF;
EXECUTE;
RUNTIME;
BKCOLOR = 3;

ACTIONS
COLOR = 0
DISPLAY
"-----"
-----"
COLOR = 4
DISPLAY "RELEVANT RANGE AND THE ESTIMATING POINT"
COLOR = 0
DISPLAY
"-----"
-----"
DISPLAY "The question of relevant range of the data must be
looked at for
each cost driver. The relevant range is usually determined by
the endpoints of the data for each cost driver but not always.
Outliers may mislead you into believing the relevant range
extends
further than is actually the case."
COLOR = 1
DISPLAY "Also, consider that you can extend past the endpoints
to some degree.
This depends on the how much confidence you have that the true
function
will not deviate much from your extrapolation."
COLOR = 4
DISPLAY "                                With this in mind --->"
COLOR = 0
LOADFACTS SETSIZE

MENU Equation_Var_Rel_Range, ALL, EQUATION, Driver
FIND Equation_Var_Rel_Range
COUNT Equation_Var_Rel_Range, COUNT_Equation_Var_Rel_Range

SAVEFACTS RELEVANT

WHILEKNOWN      POPPER_Equation_Var_Rel_Range
                RESET  POPPER_Equation_Var_Rel_Range
                POP
Equation_Var_Rel_Range, POPPER_Equation_Var_Rel_Range
                GET    POPPER_Equation_Var_Rel_Range =
Driver, EQUATION, ALL
                                Pt_In_Rng = YES
                PUT    EQUATION
```

CLOSE EQUATION  
END;

!-----RULES  
BLOCK-----

!----STATEMENTS BLOCK-----RELEVANT  
RANGE-----

ASK Equation\_Var\_Rel\_Range:  
"Look at all {DATA\_Number\_of\_Points} data points for each of  
the equation  
variables listed below. Select the variables for which the  
system you are  
estimating appears to be in the relevant range."  
;

ASK CONTINUE:" TO CONTINUE";  
CHOICES CONTINUE:\_\_\_\_PRESS\_RETURN\_\_\_\_;

PLURAL:Equation\_Var\_Rel\_Range;

# HOMOGENE.KBS

```
!-----ACTIONS BLOCK
HOMOGENEITY-----
RUNTIME;
EXECUTE;
ENDOFF;
BKCOLOR = 3;

ACTIONS
COLOR = 0
DISPLAY"-----"
-----"
COLOR = 4
DISPLAY"                                HOMOGENEITY"
COLOR = 0
DISPLAY"-----"
-----"
DISPLAY"This section applies only when you are applying this
model to"
DISPLAY"a specific system and you have obtained the necessary
data"
DISPLAY"Here we look a little bit closer at the range of the
data. Each"
DISPLAY"data point has one value for each equation variable.
The data"
DISPLAY"point you want to estimate also has a value for each
equation"
DISPLAY"variable or COST DRIVER (CD)."
```

DISPLAY"			
DISPLAY"	equation were the F-15,F-16, and		
F-111,..."			
DISPLAY"	Then you could organize the information		
as follows"			
DISPLAY"			
COLOR = 4			
DISPLAY"	CD 1 VALUE	CD 2 VALUE	CD 3
VALUE"			
COLOR = 1			
DISPLAY" OLD SYSTEMS	F-15	2	1"
DISPLAY"	F-16	2	2"
DISPLAY"	F-111	2	3"
COLOR = 0			
DISPLAY"The range for CD 1 is from 2 to 2. The range for CD 2			
is 1 to 3."			
DISPLAY"			
DISPLAY"PRESS A KEY WHEN YOU HAVE FINISHED READING ~"			

```
LOOP_RANGE = 1
LOADFACTS EQUATION
```

```

WHILE TRUE LOOP_RANGE <= (COUNT_Equation_Variables)
THEN
    LOOP_RANGE = (LOOP_RANGE + 1)
    GET System_Type = Type,EQUATION,ALL
CLS
DISPLAY"-----"
DISPLAY"-----"
COLOR = 4
DISPLAY"Look at the value of {Driver} across all the OLD DATA
POINTS"
COLOR = 0
DISPLAY"-----"
DISPLAY"-----"
    FIND          OLD_SYSTEMS_RANGE_Top
    HIGH_RANGE = (OLD_SYSTEMS_RANGE_Top)
    PUT EQUATION
    FIND          OLD_SYSTEMS_RANGE_Bottom
    LOW_RANGE = (OLD_SYSTEMS_RANGE_Bottom)
    PUT EQUATION

    OLD_SYSTEMS_RANGE_TOTAL =
((OLD_SYSTEMS_RANGE_Top)-
(OLD_SYSTEMS_RANGE_Bottom))
CLS
COLOR = 4
DISPLAY"                                {Driver}      CD VALUE      CD
VALUE"
COLOR = 0
DISPLAY"  OLD SYSTEMS"
DISPLAY"      Top of Range              {OLD_SYSTEMS_RANGE_Top}"
DISPLAY"  Bottom of Range              {OLD_SYSTEMS_RANGE_Bottom}"
DISPLAY""
DISPLAY"Find the value of {Driver} for the New System or
estimating point."
COLOR = 4
DISPLAY"  NEW SYSTEMS                    {Driver}"
DISPLAY"                                F-X1      ?"
DISPLAY""
COLOR = 0
    FIND  NEW_SYSTEMS_VALUE
DISPLAY""
COLOR = 14
    FIND  CONCLUSION_HOMOGENEITY
COLOR = 0
DISPLAY""
DISPLAY"PRESS A KEY TO CONTINUE~"

    RESET OLD_SYSTEMS_RANGE_Top
    RESET OLD_SYSTEMS_RANGE_Bottom
    RESET NEW_SYSTEMS_VALUE
    RESET CONCLUSION_HOMOGENEITY

END

```



CLOSE EQUATION  
CLS

SAVEFACTS HOMOGENE;  
!-----RULES BLOCK-----PART  
4-----

```
RULE OLD_SYSTEMS_1
IF OLD_SYSTEMS_RANGE_TOTAL = 0 AND
   NEW_SYSTEMS_VALUE       = (OLD_SYSTEMS_RANGE_Top)
THEN
CONCLUSION_HOMOGENEITY = 1
COLOR = 0
DISPLAY "THE VALUES FOR ALL THE OLD SYSTEMS AS WELL AS THE NEW
SYSTEM IS"
DISPLAY ""
COLOR = 4
DISPLAY"                                {OLD_SYSTEMS_RANGE_Top}"
DISPLAY ""
COLOR = 0
DISPLAY "No cost driver is required in this case - the
variable does not"
DISPLAY " capture any change in cost due to a change in the CD
because"
DISPLAY "                the CD is CONSTANT FOR OLD AND NEW SYSTEMS."
COLOR = 20
DISPLAY "This variable shouldn't make it to here. It can not
show up as a"
DISPLAY "                significant cost driver. Look at the data
again."
COLOR = 0 ;
```

```
RULE OLD_SYSTEMS_2
IF NEW_SYSTEMS_VALUE      >= (OLD_SYSTEMS_RANGE_Bottom) AND
   NEW_SYSTEMS_VALUE      <= (OLD_SYSTEMS_RANGE_Top)      AND
   OLD_SYSTEMS_RANGE_TOTAL > 0
THEN
CONCLUSION_HOMOGENEITY = 2
DISPLAY "In the RELEVANT RANGE. This is the ideal situation.
Be aware that although the data point is in the relevant
range, it may
still vary greatly from the data set with respect to cost.";
```

```
RULE OLD_SYSTEMS_3
IF OLD_SYSTEMS_RANGE_TOTAL <> 0 AND
   NEW_SYSTEMS_VALUE < (OLD_SYSTEMS_RANGE_Bottom) OR
   NEW_SYSTEMS_VALUE > (OLD_SYSTEMS_RANGE_Top)
THEN
CONCLUSION_HOMOGENEITY = 3
COLOR = 4
DISPLAY "Out of the relevant range!"
COLOR = 0
DISPLAY "You cannot extend to far past the relevant data range
```

```

without"
DISPLAY "          increasing the potential for estimating
error."
DISPLAY "The further you extend outside the relevant range,
the less"
DISPLAY "certainty you have that your equation will hold the
functional"
DISPLAY "          relationship.";

```

```

RULE OLD_SYSTEMS_4
IF OLD_SYSTEMS_RANGE_TOTAL = 0                                AND
   NEW_SYSTEMS_VALUE       > (OLD_SYSTEMS_RANGE_Top) OR
   NEW_SYSTEMS_VALUE       < (OLD_SYSTEMS_RANGE_Bottom)
THEN
CONCLUSION_HOMOGENEITY = 4
DISPLAY "Old Systems Same - New system Different"
COLOR = 4
DISPLAY "You cannot measure the influence of the change
because"
DISPLAY "this cost driver is constant for the old systems"
COLOR = 0 ;

```

```

!STATEMENT

```

```

BLOCK-----HOMOGENEITY-----
----
```

```

      ASK      OLD_SYSTEMS_RANGE_Top :
"What is the UPPER or TOP value of the range?  This is the
larger number in
absolute terms.";

```

```

      ASK      OLD_SYSTEMS_RANGE_Bottom :
"What is the LOWER or BOTTOM value of the range?  This is the
smaller number
in absolute terms.";

```

```

      ASK      NEW_SYSTEMS_VALUE :
"For the new system F-X1, what is the value of {Driver}?";

```

OUTWRTX.KBS

!ACTIONS BLOCK-----OUTLIERS WRT  
X-----

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS  
BCALL COPYTHREE  
COLOR = 0  
DISPLAY

"-----  
-----"

COLOR = 4  
DISPLAY "OUTLIERS WITH RESPECT TO THE X AXIS"  
COLOR = 0  
DISPLAY  
"-----  
-----"

DISPLAY"  
Potential outliers WITH RESPECT TO the X axis(WRT X) can be  
identified  
by looking at a graph of data points for each variable. There  
are four  
possibilities:

1. An outlier WRT X manifests itself as an extreme point
  2. An outlier WRT X is grouped with other points creating gaps  
in the data
  3. Both cases exist
  4. Neither case exists
- "

FIND CONTINUE  
RESET CONTINUE

LOADFACTS EQUATION  
LOOP\_WRTX = 1  
WHILETRUE COUNT\_Equation\_Variables >= (LOOP\_WRTX)  
THEN  
    LOOP\_WRTX = (LOOP\_WRTX +1)  
    POP  
Equation\_Variable\_Input, POPPER\_Equation\_Variable\_Input  
    CLS  
    COLOR = 0  
    FIND DATA\_Outliers\_WRT\_X  
    DISPLAY ""  
FIND CONTINUE  
RESET CONTINUE

```

        RESET DATA_Outliers_WRT_X
        RESET TRIGGER_DATA_Outliers_WRT_X
        RESET Extreme_Point_WRT_X
        RESET Outlier_Reason
        RESET No_Reason
    CLS
END

SAVEFACTS OUTWRTX;

!-----RULES BLOCK-----OUTLIERS WRT
X-----

    RULE DATA_OUTLIERS_1
    IF TRIGGER_DATA_Outliers_WRT_X = EXTREME_POINT
    THEN
        DATA_Outliers_WRT_X = YES
        FIND Extreme_Point_WRT_X
        FIND Outlier_Reason
        FIND No_Reason
            Driver      = (POPPER_Equation_Variable_Input)
            DataPoint   = (Extreme_Point_WRT_X)
            Reason      = (Outlier_Reason)
        APPEND OUTWRTX;

    RULE OUTLIER_REASON
    IF Outlier_Reason = Can_not_Determine
    THEN No_Reason = YES
COLOR = 4
DISPLAY"
This model can be highly influenced by the
{Extreme_Point_WRT_X}
because it is an outlier with respect to
{POPPER_EQUATION_Variable_Input}.
It appears to be a legitimate member of the population
but we don't know if there is any measurement error or not.
"
        ELSE No_Reason = NO;

    RULE DATA_OUTLIERS_2
    IF TRIGGER_DATA_Outliers_WRT_X = GAPS
    THEN
        DATA_Outliers_WRT_X = YES
        COLOR = 4
            DISPLAY "
                                THE EFFECT OF GAPS"
        COLOR = 0
            DISPLAY "When Gaps appear in the data set, the
behavior between"
            DISPLAY "the data point groupings is uncertain."
            DISPLAY "A masking effect may be taking place
which introduces"
            DISPLAY "additional potential for estimating

```

```

errors."
      DISPLAY "If the points you are estimating fall in
the Gaps,"
      DISPLAY ""
      DISPLAY "THE POTENTIAL FOR ESTIMATING ERROR IS
EXTREMELY HIGH";

      RULE DATA_OUTLIERS_3
      IF TRIGGER_DATA_Outliers_WRT_X = NEITHER
      THEN
          DATA_Outliers_WRT_X = NO
          DISPLAY "There are no Gaps or extreme points in the
data set"
          DISPLAY "This is a positive indication of model
integrity.";

      RULE DATA_OUTLIERS_4
      IF TRIGGER_DATA_Outliers_WRT_X = BOTH
      THEN
          DATA_Outliers_WRT_X = YES
          FIND Extreme_Point_WRT_X
          FIND Outlier_Reason
          FIND No_Reason
          Driver      = (POPPER_Equation_Variable_Input)
          DataPoint   = (Extreme_Point_WRT_X)
          Reason      = (Outlier_Reason)

      APPEND OUTWRTX
      CLS
      COLOR = 4
      DISPLAY "
                                     THE EFFECT OF GAPS"
      COLOR = 0
      DISPLAY "When Gaps appear in the data set, the
behavior between"
      DISPLAY "the data point groupings is uncertain."
      DISPLAY "A masking effect may be taking place which
introduces"
      DISPLAY "additional potential for estimating
errors."
      DISPLAY "If the points you are estimating fall in
the Gaps,"
      DISPLAY ""
      DISPLAY "THE POTENTIAL FOR ESTIMATING ERROR IS
EXTREMELY HIGH";

!-----STATEMENTS BLOCK-----OUTLIERS
WRTX-----

ASK TRIGGER_DATA_Outliers_WRT_X:
"
CONSIDER THE COST DRIVER: {POPPER_Equation_Variable_Input}

What situation seems to exist for this variable by looking at

```

the graph?  
";

CHOICES TRIGGER\_DATA\_Outliers\_WRT\_X:  
EXTREME\_POINT,GAPS,BOTH,NEITHER;

ASK           Extreme\_Point\_WRT\_X:  
"

EXTREME POINT LABELING

What is the name of the extreme point or suspected outlier?";

ASK Outlier\_Reason:

"You will have to determine why this point is an outlier. It may not be a legitimate member of the population. The reason may be due to measurement error, but this may not be evident to you in the write up. Pick one of the choices below:  
";

CHOICES Outlier\_Reason:Not\_Legitimate,  
                                  Measurement\_Error,  
                                  Can\_not\_Determine;

ASK CONTINUE:       "  
          TO CONTINUE";

CHOICES CONTINUE:\_\_\_PRESS\_RETURN\_\_\_;

PLURAL:Equation\_Variable\_Input;

COPYTHREE.BAT

@ ECHO OFF

COPY OUTEMPTY.DBF OUTWRTX.DBF

VARINFO.KBS

!ACTION BLOCK-----EQUATION VARIABLE  
INFORMATION-----

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS

COLOR = 0  
DISPLAY"-----  
-----"  
COLOR = 4  
DISPLAY"RATE OF CHANGE OF TECHNOLOGY"  
COLOR = 0  
DISPLAY"-----  
-----"  
DISPLAY"The rate of change of technology has to do with  
evolutionary changes"  
DISPLAY" in the methods and practices in constructing  
{SYSTEM\_TYPE}s."  
DISPLAY" For example, Aircraft avionics changed from  
simple to"  
DISPLAY"complex. This change indicates an evolutionary change  
in technology"  
DISPLAY" and may affect other cost drivers. A quantum leap in  
technology"  
DISPLAY" can not be captured by any cost driver and  
invalidates the model."  
DISPLAY" If the equation was built from data on simple  
avionics airplanes,"  
DISPLAY" and you are estimating a complex avionics  
aircraft,"  
COLOR = 4  
DISPLAY" THE RATE OF CHANGE IN TECHNOLOGY IS A FACTOR"  
COLOR = 0  
DISPLAY" Wooden pencils on the other hand are a different  
story"  
DISPLAY" because the technology is the same as when they were  
invented."  
DISPLAY""  
FIND TECHNOLOGY\_CHANGE  
CLS  
  
DISPLAY  
"-----  
-----"  
COLOR = 4  
DISPLAY "VARIABLE AND EQUATION INFORMATION"  
COLOR = 0



```

DISPLAY
"-----"
-----"
DISPLAY "The model assumes none of the variables change
direction in the"
DISPLAY "relevant range. The signs of the first and second
derivatives"
DISPLAY "must be constant throughout the estimating range."
DISPLAY ""
DISPLAY ""
DISPLAY "This portion of the program will enable you to add
information"
DISPLAY "to the database concerning the variables. Answer the
following"
DISPLAY "questions as best you can."
DISPLAY ""
DISPLAY "Menu choices are used for all questions to allow for"
DISPLAY "uniformity of responses."
DISPLAY ""
DISPLAY
"-----"
-----"
FIND CONTINUE
RESET CONTINUE
CLS

!-----FACTS-----NEW
VARIABLES

SYSTEM_NEW_VAR = NO

WHILEKNOWN RECORD_NUM
RESET RECORD_NUM
GET SYSTEM_NEW_VAR = (NEW_VAR),EQUATION,Driver
EXPERT_DRIVERS = (DRIVER)
END
COUNT EXPERT_DRIVERS,COUNT_EXPERT_DRIVERS
CLOSE EQUATION

      LOOP_Variables = 1
WHILETRUE LOOP_Variables <= (COUNT_EXPERT_DRIVERS)
  THEN
      LOOP_Variables = (LOOP_Variables +1)
  GET SYSTEM_NEW_VAR = (NEW_VAR),EQUATION,ALL
  COLOR = 4
  DISPLAY "CURRENT VARIABLE = {Driver}"
  COLOR = 0

      FIND EQUATION_VARIABLES_Character
      FIND EQUATION_VARIABLES_Sign
      PUT EQUATION
      GET EQUATION_VARIABLES_Character = (Character) AND
      EQUATION_VARIABLES_Sign = (Sign),

```

```

DERIVE,ALL
    DERIVE1 = (DERIVE1)
    DERIVE2 = (DERIVE2)
    PUT EQUATION
    FIND OTHER_FORM
    RESET OTHER_FORM

    RESET DERIVE1
    RESET DERIVE2
    RESET                                Character
    RESET    EQUATION_VARIABLES_Character
    RESET                                Sign
    RESET    EQUATION_VARIABLES_Sign

    DISPLAY"A composite variable is a single cost driver
composed of"
    DISPLAY"                several variables. For example:"
    COLOR = 4
    DISPLAY"                DENSITY = WEIGHT/VOLUME"
    COLOR = 0
    DISPLAY"                where DENSITY is a composite."
    DISPLAY""
    FIND    EQUATION_VARIABLES_Composite
    Composite=(EQUATION_VARIABLES_Composite)
    PUT    EQUATION
    RESET                                Composite
    RESET    EQUATION_VARIABLES_Composite

    FIND    EQUATION_VARIABLES_Stepindc
    Stepindc =(EQUATION_VARIABLES_Stepindc)
    PUT    EQUATION
    RESET                                Stepindc
    RESET    EQUATION_VARIABLES_Stepindc
    CLS
END
CLOSE DERIVE
CLOSE EQUATION

SYSTEM_NEW_VAR = YES
WHILEKNOWN RECORD_NUM
    RESET    RECORD_NUM
    GET    SYSTEM_NEW_VAR = (NEW_VAR),EQUATION,DRIVER
    CONTRACTOR_DRIVERS = (DRIVER)
END
COUNT CONTRACTOR_DRIVERS,COUNT_CONTRACTOR_DRIVERS
CLOSE EQUATION

    LOOP_Variables = 1
WHILETRUE    LOOP_Variables <= (COUNT_CONTRACTOR_DRIVERS)
    THEN
        LOOP_Variables = (LOOP_Variables +1)
    GET    SYSTEM_NEW_VAR = (NEW_VAR),EQUATION,ALL
    COLOR = 4

```

```

DISPLAY "CURRENT VARIABLE = {Driver}"
COLOR = 0

MENU      EQUATION_VARIABLES_Factor,ALL,YAKBASE,Factor
FIND      EQUATION_VARIABLES_Factor
Factor =  (EQUATION_VARIABLES_Factor)
PUT       EQUATION
RESET     Factor
RESET     EQUATION_VARIABLES_Factor
MRESET    EQUATION_VARIABLES_Factor

FIND      EQUATION_VARIABLES_Character
FIND      EQUATION_VARIABLES_Sign
PUT       EQUATION
      GET  EQUATION_VARIABLES_Character = (Character) AND
      EQUATION_VARIABLES_Sign          = (Sign),

DERIVE,ALL
      DERIVE1 = (DERIVE1)
      DERIVE2 = (DERIVE2)

DISPLAY"
The sign of the First Derivative  = {DERIVE1}
The sign of the Second Derivative = {DERIVE2}"
PUT EQUATION

      FIND OTHER_FORM
      RESET OTHER_FORM

RESET     Character
RESET     EQUATION_VARIABLES_Character
RESET     Sign
RESET     EQUATION_VARIABLES_Sign

DISPLAY"A composite variable is a single cost driver
composed of"
DISPLAY"      several variables. For example:"
COLOR = 4
DISPLAY"      DENSITY = WEIGHT/VOLUME"
COLOR = 0
DISPLAY""
FIND      EQUATION_VARIABLES_Composite
Composite=(EQUATION_VARIABLES_Composite)
PUT       EQUATION
RESET     Composite
RESET     EQUATION_VARIABLES_Composite
FIND      EQUATION_VARIABLES_Stepindc
Stepindc =(EQUATION_VARIABLES_Stepindc)
PUT       EQUATION
RESET     Stepindc
RESET     EQUATION_VARIABLES_Stepindc
CLS
END
CLOSE DEFINE
CLOSE EQUATION
CLOSE YAKBASE;

```

!RULES BLOCK-----EQUATION  
INFORMATION-----

```
RULE      TECHNOLOGY_CHANGE
IF        TRIGGER_TECHNOLOGY_CHANGE = YES
THEN      TECHNOLOGY_CHANGE = YES
           TYPE      = (SYSTEM_TYPE)
           FACTOR    = CHANGE_OF_TECHNOLOGY
           APPEND YAKBASE
           DISPLAY" A cost driver may be need to capture this
factor."
           DISPLAY" PRESS ANY KEY TO CONTINUE~"
ELSE
           TECHNOLOGY_CHANGE = NO
DISPLAY"No cost driver needed to capture technology change in
this field."
DISPLAY"PRESS ANY KEY TO CONTINUE~";
```

```
RULE OTHER_FORM
IF EQUATION_VARIABLES_Character = OTHER
THEN
OTHER_FORM = YES
FIND DERIVE1
FIND DERIVE2
PUT EQUATION
RESET DERIVE1
RESET DERIVE2
ELSE
OTHER_FORM = NO;
```

!STATEMENT BLOCK-----EQUATION  
INFORMATION-----

```
ASK CONTINUE:"      TO CONTINUE";
CHOICES CONTINUE:____PRESS_RETURN____;
```

```
ASK      TRIGGER_TECHNOLOGY_CHANGE:
"Is the technology changing so rapidly in this area as to
affect the
nature of the other factors and variables?";
CHOICES TRIGGER_TECHNOLOGY_CHANGE: YES,NO;
```

```
ASK      EQUATION_VARIABLES_Factor:
"Which Factor does this cost driver capture?";
```

```
ASK      EQUATION_VARIABLES_Character:
"
What form does this variable take in the equation?
";
CHOICES
EQUATION_VARIABLES_Character:X,X_SQUARED,X_CUBED,1_OVER_X,SQ_R
OOT_OF_X,
LOG_X,1_OVER_SQT_X,OTHER;
```

ASK EQUATION\_VARIABLES\_Sign:  
"What sign does this variable have in the equation?";  
CHOICES EQUATION\_VARIABLES\_Sign:PLUS,MINUS;

ASK Derive1:  
"Looking at this variable, what is the sign of the first derivative?";  
CHOICES Derive1:POSITIVE,NEGATIVE;

ASK Derive2:  
"What is the sign of the second derivative?";  
CHOICES Derive2:POSITIVE,NEGATIVE;

ASK EQUATION\_VARIABLES\_Composite:  
"Is this variable a composite variable in the equation?";  
CHOICES EQUATION\_VARIABLES\_Composite:YES,NO;

ASK EQUATION\_VARIABLES\_Stepindc:  
"Is this an indicator variable?";  
CHOICES EQUATION\_VARIABLES\_Stepindc:YES,NO;

PLURAL: EQUATION\_VARIABLES,  
EXPERT\_DRIVERS,  
CONTRACTOR\_DRIVERS;  
AUTOQUERY;

LISTCON.KBS

!ACTIONS BLOCK-----CONCLUSIONS FOR  
TYPELIST-----  
ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

ACTIONS  
COLOR = 1

LOADFACTS TYPELIST

COLOR = 4  
DISPLAY "-----CONCLUSIONS FROM COST DRIVER  
CONSIDERATIONS-----"  
DISPLAY ""  
COLOR = 0  
DISPLAY ""  
DISPLAY "The {System\_Type}s database is used for this  
consultation."  
COLOR = 4  
DISPLAY ""  
DISPLAY "COST DRIVERS FROM THE EXPERT'S LIST AND CONTRACTOR'S  
LIST"  
COLOR = 0  
DISPLAY ""

FIND DISPLAY\_Listed\_Factors

CLS

COLOR = 0  
DISPLAY "-----  
-----"  
COLOR = 4  
DISPLAY "COST DRIVERS FROM THE CONTRACTOR'S LIST  
ONLY"  
COLOR = 0  
DISPLAY "-----  
-----"

FIND DISPLAY\_More\_Listed\_Factors

DISPLAY "-----  
-----"  
SAVEFACTS LISTCON  
CLS;

!RULES BLOCK-----CONCLUSIONS FOR  
TYPELIST-----

RULE DISPLAY\_EXPERT\_AND\_CONTRACTOR  
IF COUNT\_Listed\_Factors > 0  
THEN DISPLAY\_Listed\_Factors = YES

DISPLAY "The cost drivers in the contractor list AND in the expert list are:"

COLOR = 1

CLOSE YAKBASE

RESET RECORD\_NUM

WHILEKNOWN RECORD\_NUM

RESET RECORD\_NUM

SET\_In\_List = YES

SET\_New\_Var = NO

GET SET\_In\_List = In\_List AND

SET\_New\_Var = New\_Var, YAKBASE, Driver

DISPLAY "{Driver}"

END

CLOSE YAKBASE

COLOR = 0

FIND CONTINUE

RESET CONTINUE

FIND Excluded\_Factors

ELSE DISPLAY\_Listed\_Factors = NO;

RULE EXCLUDED

IF MENU\_SIZE = (COUNT\_LISTED\_FACTORS)

THEN

Excluded\_Factors = NONE

CLS

COLOR = 0

DISPLAY"-----"

COLOR = 1

DISPLAY"No cost drivers listed by the experts were excluded.

This indicates"

DISPLAY" that the contractor considered all the relevant cost drivers."

COLOR = 0

DISPLAY"-----"

FIND CONTINUE

RESET CONTINUE

ELSE

Excluded\_Factors = SOME

CLS

COLOR = 4

DISPLAY" -----CAUTION - EXCLUDED COST DRIVERS"

COLOR = 0

DISPLAY" The following cost drivers were"

DISPLAY" listed in the expert database"

DISPLAY" but NOT listed in the contractors list:"

COLOR = 4

CLOSE YAKBASE

RESET RECORD\_NUM

WHILEKNOWN RECORD\_NUM

RESET RECORD\_NUM

```

        SET_In_List = NO
        SET_New_Var = NO
    GET SET_In_List = In_List AND
        SET_New_Var = New_Var AND
        SYSTEM_Type = Type ,YAKBASE,Driver
        DISPLAY "{Driver}"
        LEFT_OUT_VARIABLES = (Driver)
    END
    CLOSE YAKBASE
    COUNT LEFT_OUT_VARIABLES,COUNT_NAMES
COLOR = 0
DISPLAY"The contractor should have provided a reason for
excluding any"
DISPLAY"variables. The experts who built the database see all
the cost"
DISPLAY"drivers as important factors for consideration."
DISPLAY""
FIND CONTINUE
RESET CONTINUE
FIND LEFT_OUT_REASONS;

RULE LEFT_OUT_REASONS
IF Excluded_Factors = SOME
THEN
LEFT_OUT_REASONS = WHY
LOOP_LEFT_OUT = 1
WHILETRUE COUNT_NAMES >= (LOOP_LEFT_OUT)
THEN
LOOP_LEFT_OUT = (LOOP_LEFT_OUT +1)
POP LEFT_OUT_VARIABLES,LEFT_OUT_NAME
CLS
FIND WHY_LEFT_OUT
RESET WHY_LEFT_OUT
RESET TRIGGER_WHY_LEFT_OUT
DISPLAY"
"
FIND CONTINUE
RESET CONTINUE
END;

RULE EXCLUDED_CD_1
IF TRIGGER_WHY_LEFT_OUT = 1
THEN WHY_LEFT_OUT = 1
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : ALTERNATE MEASURE IN THE MODEL"
COLOR = 0
DISPLAY"This is an acceptable reason, NO increase in risk";

RULE EXCLUDED_CD_2
IF TRIGGER_WHY_LEFT_OUT = 2

```



```

THEN          WHY_LEFT_OUT = 2
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : DATA WAS NOT KNOWABLE OR
MEASURABLE"
COLOR = 0
DISPLAY"Although this is an acceptable reason, this constrains
the model.";

RULE EXCLUDED_CD_3
IF TRIGGER_WHY_LEFT_OUT = 3
THEN          WHY_LEFT_OUT = 3
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : DATA WAS NOT AVAILABLE"
COLOR = 0
DISPLAY"Although this is an acceptable reason, this constrains
the model.";

RULE EXCLUDED_CD_4
IF TRIGGER_WHY_LEFT_OUT = 4
THEN          WHY_LEFT_OUT = 4
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : COLLINEARITY"
COLOR = 0
DISPLAY"Alternate models should be developed to explore this
variable."
DISPLAY"Any time variables are omitted from the equation"
DISPLAY"collinearity may be a suspected problem."
DISPLAY"If the cost driver was omitted from the equation to
eliminate"
DISPLAY"collinearity problems, the equation may not be
capturing"
DISPLAY"all the factor that the variable represented. If the
collinear"
DISPLAY"relationship between the omitted variable and the
equation variable(s)"
DISPLAY"holds for the new estimate points, the equation is
acceptable. If, on"
DISPLAY"the other hand, the relationship between the omitted
variable and"
DISPLAY"the equation variable(s) changes for the new
estimating point,"
COLOR = 4
DISPLAY"the problem will not shown up as wide confidence
bouds but"

```

```

DISPLAY"CONSIDERABLE ESTIMATING ERRORS MAY OCCUR."
COLOR = 0;

RULE EXCLUDED_CD_5
IF TRIGGER_WHY_LEFT_OUT = 5
THEN WHY_LEFT_OUT = 5
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : INSIGNIFICANT IN COMBINATION WITH
OTHER VARIABLES"
COLOR = 0
DISPLAY"This variable was dropped out due to statistical
insignificance."
DISPLAY"The model may be misspecified or misidentified. This
is addressed"
DISPLAY"later in the program. Some risk may be present.";

RULE EXCLUDED_CD_6
IF TRIGGER_WHY_LEFT_OUT = 6
THEN WHY_LEFT_OUT = 6
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : INSIGNIFICANT DUE TO THE EFFECT OF
OUTLIERS"
COLOR = 0
DISPLAY"This variable may have been significant if outliers
did not exist
hence some amount of cost variation may be lost. INCREASED
RISK!";

RULE EXCLUDED_CD_7
IF TRIGGER_WHY_LEFT_OUT = 7
THEN WHY_LEFT_OUT = 7
DISPLAY""
CLS
DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : SAMPLE DOES NOT REPRESENT
POPULATION"
COLOR = 0
DISPLAY"A bad sample misrepresents the population and
increases risk.";

RULE EXCLUDED_CD_8
IF TRIGGER_WHY_LEFT_OUT = 8
THEN WHY_LEFT_OUT = 8
DISPLAY""
CLS

```

```

DISPLAY"EXCLUDED VARIABLE: {LEFT_OUT_NAME} "
COLOR = 4
DISPLAY"REASON EXCLUDED : NONE GIVEN"
COLOR = 0
DISPLAY"When no reason is given for exclusion, risk is can be
extremely high.";

```

```

RULE DISPLAY_CONTRACTOR_ONLY
IF      More_Listed_Factors = YES
THEN DISPLAY_More_Listed_Factors = YES
    DISPLAY "The cost driver(s) which you allowed that were
listed in the"
    DISPLAY "contractor list AND NOT listed in the expert list
are:"
    DISPLAY ""
    COLOR = 1
        CLOSE YAKBASE
        RESET      RECORD_NUM
        WHILEKNOWN RECORD_NUM
        RESET      RECORD_NUM
        SET_In_List = YES
        SET_New_Var = YES
    GET SET_In_List = In_List AND
    SET_New_Var = New_Var, YAKBASE,Driver
    DISPLAY "{Driver}"
    END
    CLOSE YAKBASE
    COLOR = 0
    FIND CONTINUE
    RESET CONTINUE
    DISPLAY "The cost driver(s) which you DID NOT allowed that
were listed in the"
    DISPLAY "contractor list AND NOT listed in the expert list
are:"
    DISPLAY ""
    COLOR = 4
        CLOSE YAKBASE
        RESET      RECORD_NUM
        WHILEKNOWN RECORD_NUM
        RESET      RECORD_NUM
        SET_In_List = NO
        SET_New_Var = YES
    GET SET_In_List = In_List AND
    SET_New_Var = New_Var, YAKBASE,Driver
    DISPLAY "{Driver}"
    END
    CLOSE YAKBASE
    COLOR = 0
    DISPLAY"Any variables listed here were not considered valid
for any"
    DISPLAY"use in this model."
    FIND CONTINUE

```

```

RESET CONTINUE
ELSE More_Listed_Factors = NO
DISPLAY "No cost drivers are listed by the contractor only."
DISPLAY ""
FIND CONTINUE
RESET CONTINUE;

```

```

!-----STATEMENTS BLOCK-----LIST
CONCLUSIONS-----

```

```

ASK CONTINUE:"      TO CONTINUE";
CHOICES CONTINUE:___PRESS_RETURN___;

```

```

ASK      SWITCH_Definition:
"Choose ? if you are done or EDIT to change the definition.";

```

```

ASK      TRIGGER_WHY_LEFT_OUT:
"Why was {LEFT_OUT_NAME} left out?
1. There is an alternate measure (Ex. two different measures
of weight).
2. The information for this variable was not obtainable (not
measurable).
3. No data available for this cost driver (measurable but not
available).
-- This variable or cost driver was statistically
insignificant when
   brought into the equation. This indicated
4. Collinearity became a problem when this variable was
brought in.
5. The variable is insignificant when combined with certain
other variables.
6. The variable is insignificant due to the effects of an
outlier.
7. This sample does not represent the population.
8. No reason given
";

```

```

CHOICES TRIGGER_WHY_LEFT_OUT:1,2,3,4,5,6,7,8;

```

```

PLURAL: HOLDER_More_Listed_Factors,
        Type,
        Listed_Factors,
        Left_Out_Variables;

```

RELEVCON.KBS

!ACTIONS BLOCK-----CONCLUSION RELEVANT

RANGE-----

ENDOFF;

EXECUTE;

RUNTIME;

BKCOLOR = 3;

ACTIONS

COLOR = 0

DISPLAY

"-----"

-----"

COLOR = 4

DISPLAY " CONCLUSIONS ON THE RELEVANT RANGE AND THE  
ESTIMATING POINT"

COLOR = 0

DISPLAY

"-----"

-----"

DISPLAY ""

LOADFACTS EQUATION

LOADFACTS RELEVANT

DIFFERENCE

=((COUNT\_Equation\_Variables)-(COUNT\_Equation\_Var\_Rel\_Range))

FIND CONCLUDE\_Equation\_Var\_Rel\_Range

SAVEFACTS RELEVCON;

!-----RULES

BLOCK-----

RULE CONCLUSION\_RELEVANT\_RANGE

IF COUNT\_Equation\_Var\_Rel\_Range =

(COUNT\_Equation\_Variables)

THEN CONCLUDE\_Equation\_Var\_Rel\_Range = Good

DISPLAY ""

COLOR = 1

DISPLAY "Your point estimate is in the Relevant  
Range for all cost"

DISPLAY "drivers. This is a positive indication of  
model integrity."

DISPLAY ""

FIND CONTINUE

RESET CONTINUE

ELSE CONCLUDE\_Equation\_Var\_Rel\_Range = Bad

COLOR = 4

DISPLAY "Number of Variables not in relevant range:  
{DIFFERENCE}"

```
COLOR = 0
DISPLAY""
DISPLAY"When your estimate has values outside of the relevant
range,"
DISPLAY"for any cost driver, the model behavior is
unpredictable."
DISPLAY""
COLOR = 4
DISPLAY"THIS IS AN INDICATION OF HIGH POTENTIAL ESTIMATING
ERROR"
DISPLAY""
COLOR = 0
FIND CONTINUE
RESET CONTINUE;
```

```
!-----ASK STATEMENTS AND RELEVANT RANGE
CONCLUSIONS-----
```

```
ASK CONTINUE:"      TO CONTINUE";
CHOICES CONTINUE:___PRESS_RETURN___;
```

```
PLURAL:Equation_Var_Rel_Range;
```

OVERCON.KBS

!ACTIONS BLOCK-----CONCLUSION OVERLOOKED  
DATAPOINTS-----  
ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;  
ACTIONS  
COLOR = 4

LOADFACTS OVERLOOK  
  DISPLAY ""  
  DISPLAY "----CONCLUSIONS: OVERLOOKED MEMBERS OF THE DEFINED  
POPULATION----"  
  COLOR = 0  
  DISPLAY""  
  FIND OVERLOOK  
  DISPLAY ""  
  FIND CONTINUE  
  RESET CONTINUE;

!RULES BLOCK-----CONCLUSION  
OVERLOOKS-----

RULE OVERLOOK  
  IF TRIGGER\_Overlook = Yes  
  THEN OVERLOOK = YES  
    DISPLAY "This cost model was based on a population  
which did"  
    DISPLAY "not include all the members possible."  
    DISPLAY "This can decrease the integrity of the  
regression process"  
    DISPLAY "if the sample used was not chosen randomly."  
  ELSE OVERLOOK= NO  
    DISPLAY "This cost model was based on a complete  
dataset."  
    DISPLAY "This is a positive indication of model  
integrity.";

!STATEMENTS BLOCK-----CONCLUSION  
OVERLOOKS-----

ASK CONTINUE:"          TO CONTINUE";  
CHOICES CONTINUE:\_\_\_\_PRESS\_RETURN\_\_\_\_;

# FACTRCON.KBS

```
!ACTION
BLOCK-----FACTRCON-----
-----
```

```
ENDOFF;
EXECUTE;
RUNTIME;
BKCOLOR = 3;
```

## ACTIONS

```
COLOR = 0
DISPLAY"-----
-----"
COLOR = 4
DISPLAY "CONCLUSIONS ON CAPTURING THE KEY FACTORS WITH
EQUATION VARIABLES"
COLOR = 0
DISPLAY"-----
-----"
```

## LOADFACTS EQUATION

```
WHILEKNOWN RECORD_NUM
    RESET RECORD_NUM
    GET SYSTEM_TYPE = TYPE,AYAKBASE,FACTOR
    HOLDER_DBASE_FACTORS = (FACTOR)
END
COUNT HOLDER_DBASE_FACTORS,COUNT_DBASE_FACTORS
CLOSE YAKBASE
```

```
DISPLAY"The experts have identified {COUNT_DBASE_FACTORS}
factors that"
DISPLAY"should be represented by variables in the equation."
COLOR = 1
DISPLAY"THEY ARE:"
DISPLAY"{HOLDER_DBASE_FACTORS}"
COLOR = 0
DISPLAY""
DISPLAY"Each variable captures one of the factors listed
above. Ideally, there"
DISPLAY"          should be one at least one variable per
factor."
DISPLAY"          If there are no variables, the factor is not
addressed."
DISPLAY""
FIND CONTINUE
RESET CONTINUE
CLS
```

```
    LOOP_DBASE_FACTORS = 1
```



```

WHILE TRUE LOOP_DBASE_FACTORS <= (COUNT_DBASE_FACTORS)
THEN
    LOOP_DBASE_FACTORS = (LOOP_DBASE_FACTORS + 1)
POP    HOLDER_DBASE_FACTORS,    POPPER_DBASE_FACTORS
GET    POPPER_DBASE_FACTORS = (FACTOR), EQUATION, ALL
        HOLDER_SAME_FACTOR_EQUATION_VARIABLE = (DRIVER)
COUNT HOLDER_SAME_FACTOR_EQUATION_VARIABLE,
        COUNT_SAME_FACTOR_EQUATION_VARIABLE
COLOR = 0
DISPLAY
"-----"
"-----"
COLOR = 4
DISPLAY "ANALYZING THE FACTOR = {POPPER_DBASE_FACTORS}"
COLOR = 0
DISPLAY
"-----"
"-----"
FIND    ZERO_DRIVERS
FIND    ONE_DRIVER
FIND    TWO_DRIVERS
FIND    THREE_DRIVERS
CLOSE EQUATION
RESET    ZERO_DRIVERS
RESET    ONE_DRIVER
RESET    TWO_DRIVERS
RESET    THREE_DRIVERS
RESET    HOLDER_SAME_FACTOR_EQUATION_VARIABLE
RESET    COUNT_SAME_FACTOR_EQUATION_VARIABLE
FIND CONTINUE
RESET CONTINUE
CLS
END ;

!-----RULES BLOCK-----PART
6-----
RULE ZERO_DRIVERS
IF    COUNT_SAME_FACTOR_EQUATION_VARIABLE = 0
THEN
    ZERO_DRIVERS=YES
    ONE_DRIVER  =NO
    TWO_DRIVERS =NO
THREE_DRIVERS =NO
    DISPLAY "THIS ARE NO VARIABLES ASSOCIATED WITH
{POPPER_DBASE_FACTORS}."
    DISPLAY ""
    COLOR = 20
    DISPLAY "THE EFFECT ON COST THIS FACTOR IS HAVING IS NOT
BEING CAPTURED."
    COLOR = 0
    DISPLAY ""
    DISPLAY "This may be a problem.  The factors for this system
were picked"

```

```

    DISPLAY "by experts in the field.  If a factor is not
    represented by a "
    DISPLAY "cost driver or equation variable, the equation
    challenges the"
    DISPLAY "database built by the experts in this field.  This
    may lead to"
    DISPLAY "an increased potential for estimating error."
    DISPLAY""
ELSE ZERO_DRIVERS=NO
GET      POPPER_DBASE_FACTORS =
(FACTOR),EQUATION,ALL
        HOLDER_SAME_FACTOR_EQUATION_VARIABLE = (DRIVER)
RESET    COUNT_SAME_FACTOR_EQUATION_VARIABLE
COUNT   HOLDER_SAME_FACTOR_EQUATION_VARIABLE,
        COUNT_SAME_FACTOR_EQUATION_VARIABLE;

RULE ONE_DRIVER
IF  COUNT_SAME_FACTOR_EQUATION_VARIABLE =1
THEN
    ONE_DRIVER  =YES
    TWO_DRIVERS =NO
    THREE_DRIVERS =NO
    DISPLAY "THERE IS 1 VARIABLE ASSOCIATED WITH
    {POPPER_DBASE_FACTORS}."
    DISPLAY "IT IS: "
    DISPLAY""
    DISPLAY "{HOLDER_SAME_FACTOR_EQUATION_VARIABLE}"
    DISPLAY ""
    DISPLAY "With one cost driver, you have to make a judgement
    call."
    DISPLAY "Does this variable capture all the change in the
    factor?"
    DISPLAY "If you think it does, than this is adequate.  If not,
    the"
    DISPLAY "change in this factor is not being totally captured
    and"
    DISPLAY "this may result in an inaccurate estimate of cost."
    DISPLAY ""
ELSE ONE_DRIVER=NO
GET      POPPER_DBASE_FACTORS =
(FACTOR),EQUATION,ALL
        HOLDER_SAME_FACTOR_EQUATION_VARIABLE = (DRIVER)
RESET    COUNT_SAME_FACTOR_EQUATION_VARIABLE
COUNT   HOLDER_SAME_FACTOR_EQUATION_VARIABLE,
        COUNT_SAME_FACTOR_EQUATION_VARIABLE;

RULE TWO_DRIVERS
IF  COUNT_SAME_FACTOR_EQUATION_VARIABLE =2
THEN
    TWO_DRIVERS =YES
    THREE_DRIVERS =NO
    DISPLAY "THERE ARE 2 VARIABLES ASSOCIATED WITH
    {POPPER_DBASE_FACTORS}."

```

```

DISPLAY "THEY ARE: "
DISPLAY""
DISPLAY "{HOLDER_SAME_FACTOR_EQUATION_VARIABLE}"
DISPLAY ""
DISPLAY "Two cost drivers are usually enough to capture any
change in"
DISPLAY "a factor."
DISPLAY""
ELSE TWO_DRIVERS=NO
GET      POPPER_DBASE_FACTORS =
(FACTOR),EQUATION,ALL
      HOLDER_SAME_FACTOR_EQUATION_VARIABLE = (DRIVER)
RESET    COUNT_SAME_FACTOR_EQUATION_VARIABLE
COUNT   HOLDER_SAME_FACTOR_EQUATION_VARIABLE,
      COUNT_SAME_FACTOR_EQUATION_VARIABLE;

RULE THREE_DRIVERS
IF  COUNT_SAME_FACTOR_EQUATION_VARIABLE =3
THEN
THREE_DRIVERS =YES
DISPLAY "THERE ARE 3 VARIABLES ASSOCIATED WITH
{POPPER_DBASE_FACTORS}."
DISPLAY "THEY ARE: "
DISPLAY""
DISPLAY "{HOLDER_SAME_FACTOR_EQUATION_VARIABLE}"
DISPLAY ""
DISPLAY "Three cost drivers should capture all  the change in
a factor.";

!-----STATEMENT BLOCK-----CONCLUSIONS
FACTORS-----

ASK CONTINUE:"      TO CONTINUE";
CHOICES CONTINUE:____PRESS_RETURN____;

PLURAL: HOLDER_DBASE_FACTORS,
      HOLDER_SAME_FACTOR_EQUATION_VARIABLE;
AUTOQUERY;

```

# IDENCON.KBS

!ACTION

BLOCK-----IDENCON-----  
-----

ENDOFF;  
EXECUTE;  
RUNTIME;  
BKCOLOR = 3;

## ACTIONS

### LOADFACTS EQUATION

```
LOOP_IDEN = 1
WHILETRUE LOOP_IDEN <= (COUNT_EQUATION_VARIABLES)
THEN
  LOOP_IDEN = (LOOP_IDEN +1)
  GET ALL,EQUATION,ALL
    Equation_Driver = (Driver)
    Equation_Composite = (Composite)
    Equation_Stepindc = (Stepindc)
  RESET DRIVER
  RESET COMPOSITE
  RESET STEPINDC
  GET Equation_Driver = (Driver),YAKBASE,ALL
  RESET EXPERT_DRIVER
  RESET EXPERT_COMPOSITE
  RESET EXPERT_STEPINDC
    Expert_Driver = (Driver)
    Expert_Composite = (Composite)
    Expert_Stepindc = (Stepindc)
  COLOR = 0
  DISPLAY"-----"
  -----"
  COLOR = 4
  DISPLAY"CONCLUSION CONCERNING COMPOSITE AND INDICATOR
  VARIABLES"
  COLOR = 0
  DISPLAY"-----"
  -----"
  COLOR = 4
  DISPLAY"CURRENT VARIABLE = {DRIVER}"
  COLOR =0
  DISPLAY""
  DISPLAY"The expert database has classified this variable as
  follows:
  Is it a composite variable ---> {Expert_Composite}
  Is it an indicator variable ---> {Expert_Stepindc}"
  DISPLAY"The equation classifies it as:
  Is it a composite variable ---> {Equation_Composite}
  Is it an indicator variable ---> {Equation_Stepindc}
```

```

"
FIND MATCH_ALL
RESET MATCH_ALL
DISPLAY""
FIND CONTINUE
RESET CONTINUE

CLS
END;

!-----RULES BLOCK-----CONCLUSIONS ON COMPS AND
IND-----

RULE New_Var
IF      Expert_Composite = ? AND
      Expert_Stepindc = ?
THEN
    MATCH_ALL = NO
    DISPLAY"The experts database has no basis from which this
variable can be"
    DISPLAY"compared. The variable stands as specified by the
contractor.";

RULE ALL_MATCH
IF      Equation_Composite =(Expert_Composite) AND
      Equation_Stepindc =(Expert_Stepindc)
THEN
    MATCH_ALL = YES
    DISPLAY""
    DISPLAY"There is no conflict with the identified of this
variable."
    DISPLAY"";

RULE StepIndc_Mismatch
IF      Equation_Composite =(Expert_Composite) AND
      Equation_Stepindc <>(Expert_Stepindc)
THEN
    MATCH_ALL = NO
    DISPLAY""
    DISPLAY"The expert database and the equation variable are in
conflict"
    DISPLAY"concerning the INDICATOR variable classification."
    DISPLAY""
    COLOR = 4
    DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
    COLOR = 0;

RULE Composite_Mismatch

```

```

IF      Equation_Composite <>(Expert_Composite) AND
      Equation_Stepindc  =(Expert_Stepindc)
THEN
      MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"
DISPLAY"concerning the COMPOSITE variable classification."
DISPLAY""
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

RULE Both_Mismatch
IF      Equation_Composite <>(Expert_Composite) AND
      Equation_Stepindc  <>(Expert_Stepindc)
THEN
      MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"
DISPLAY"concerning both COMPOSITE and INDICATOR variable
classification."
DISPLAY""
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

!-----STATEMENT BLOCK-----IDENTIFICATION
CONCLUSIONS

```

```

ASK CONTINUE:"          TO CONTINUE";
CHOICES CONTINUE:____PRESS_CONTINUE____;

```

SPECCON1.KBS

```
!ACTION BLOCK-----SPECIFICATION -----
ENDOFF;
EXECUTE;
RUNTIME;
BKCOLOR = 3;
```

ACTIONS

COLOR = 4

```
DISPLAY"-----SPECIFICATION-----
-----"
```

COLOR = 0

DISPLAY"The specification part of the analysis deals with the form of the"

DISPLAY" cost drivers or variables. Most of the data for this section"

DISPLAY" was input previously."

COLOR = 4

```
DISPLAY"-----
-----"
```

COLOR = 0

DISPLAY""

FIND CONTINUE

RESET CONTINUE

CLS

COLOR = 4

```
DISPLAY"-----OUTLIERS WITH RESPECT TO
Y-----"
```

COLOR = 0

LOADFACTS EQUATION

FIND PERCENT

FIND OUTLIERS\_WRT\_Y

LOOP\_OUTLIERS = 1

WHILETRUE LOOP\_OUTLIERS <= (COUNT\_OUTLIER\_POINT)

THEN LOOP\_OUTLIERS = (LOOP\_OUTLIERS + 1)

POP HOLDER\_OUTLIER\_POINT,

POPPER\_OUTLIER\_POINT

DISPLAY""

COLOR = 4

DISPLAY"---{POPPER\_OUTLIER\_POINT}---"

COLOR = 0

DISPLAY""

FIND OUTLIER\_REASON

RESET OUTLIER\_REASON

RESET TRIGGER\_OUTLIER\_REASON

END

DISPLAY""

CLS

COLOR = 4

```

DISPLAY"-----LEVERAGE
VALUES-----"
COLOR = 0

```

```

LOADFACTS EQUATION
LOADFACTS SETSIZE

```

```

MEAN_LEV_VAL =
(2*(COUNT_EQUATION_VARIABLES)/(DATA_NUMBER_OF_POINTS))

```

```

DISPLAY"Leverage values will reveal outliers WRT X which are
not extreme"

```

```

DISPLAY"          points but weird combinations of X and Y."

```

```

DISPLAY"There are several ways to determine acceptable
leverage values. Two"

```

```

DISPLAY"criteria are: "

```

```

COLOR = 4

```

```

DISPLAY"1. A leverage value of .7 or above is considered high"

```

```

DISPLAY"2. A leverage value greater than two times the mean
leverage value"

```

```

DISPLAY"  which is {MEAN_LEV_VAL} in this case."

```

```

COLOR = 0

```

```

DISPLAY"          Look at each data point for each variable."

```

```

WHILEKNOWN SWITCH_HIGHEST_LEVERAGE_VALUE

```

```

HOLDER_HIGHEST_LEVERAGE_VALUE=(SWITCH_HIGHEST_LEVERAGE_VALUE)

```

```

  RESET    SWITCH_HIGHEST_LEVERAGE_VALUE

```

```

DISPLAY""

```

```

  FIND    SWITCH_HIGHEST_LEVERAGE_VALUE

```

```

DISPLAY""

```

```

CLS

```

```

END

```

```

COUNT    HOLDER_HIGHEST_LEVERAGE_VALUE,COUNT_LEVERAGE

```

```

      LOOP_LEVERAGE = 1

```

```

WHILETRUE LOOP_LEVERAGE <= (COUNT_LEVERAGE)

```

```

THEN    LOOP_LEVERAGE = (LOOP_LEVERAGE +1)

```

```

  POP    HOLDER_HIGHEST_LEVERAGE_VALUE,

```

```

        POPPER_HIGHEST_LEVERAGE_VALUE

```

```

DISPLAY""

```

```

COLOR =4

```

```

DISPLAY"-----{POPPER_HIGHEST_LEVERAGE_VALUE}-----"

```

```

COLOR =0

```

```

DISPLAY""

```

```

  FIND    LEVERAGE_REASON

```

```

DISPLAY""

```

```

  RESET    LEVERAGE_REASON

```

```

  RESET    TRIGGER_LEVERAGE_REASON

```

```

COLOR = 4

```

```

FIND CONTINUE

```

```

RESET CONTINUE

```

```

CLS

```

```

COLOR = 0

```



END

DISPLAY""

CLS

COLOR = 4

DISPLAY"-----NORMALITY

PLOTS-----"

COLOR = 0

DISPLAY"Normality plots plot each residual against its  
expected value when"

DISPLAY"the distribution is normal. A plot that is a 45  
degree line suggests"

DISPLAY"agreement with normality, whereas a plot that departs  
substantially"

DISPLAY"from linearity suggests that the error distribution is  
not normal."

DISPLAY""

FIND NORMALITY

DISPLAY""

FIND CONTINUE

RESET CONTINUE

DISPLAY""

CLS

COLOR = 4

DISPLAY"-----HETEROSCEDASTICITY-----

-----"

COLOR = 0

DISPLAY"Heteroscedasticity is the condition of the error  
variance not being"

DISPLAY"constant over all cases, in contrast to the condition  
of equal error"

DISPLAY"variances called HOMOSCEDASTICITY."

DISPLAY ""

FIND HETEROSCEDASTICITY

DISPLAY""

FIND CONTINUE

RESET CONTINUE

DISPLAY""

CLS

COLOR = 4

DISPLAY"-----AUTOCORRELATION-----

-----"

COLOR = 0

DISPLAY"AUTOCORRELATION is the condition of the error terms  
being correlated"

DISPLAY"over time. For this to be a factor the model, the  
data needs to have"

DISPLAY"a constant lag (time-series data)."

DISPLAY""

FIND AUTOCORRELATION

```

DISPLAY""
FIND CONTINUE
RESET CONTINUE
;

!RULES
BLOCK-----SPECIFICATION-----
-----

RULE          OUTLIERS_WRT_Y
IF TRIGGER_OUTLIERS_WRT_Y = NO
THEN          OUTLIERS_WRT_Y = NO
DISPLAY"The residual plots indicate you have no outliers with
respect to Y."
DISPLAY""
FIND CONTINUE
RESET CONTINUE
ELSE          OUTLIERS_WRT_Y = YES
              WHILEKNOWN SWITCH_OUTLIERS_WRT_Y

HOLDER_OUTLIER_POINT=(SWITCH_OUTLIERS_WRT_Y)
              RESET          SWITCH_OUTLIERS_WRT_Y
              FIND          SWITCH_OUTLIERS_WRT_Y
              END
              COUNT  HOLDER_OUTLIER_POINT,COUNT_OUTLIER_POINT;

RULE          OUTLIER_REASON_1
IF TRIGGER_OUTLIER_REASON =1
THEN          OUTLIER_REASON =1
DISPLAY"The data point {POPPER_OUTLIER_POINT} is an outlier
due to model misspecification."
DISPLAY"The only remedy for this is to fix the model,
(respecification).\"
COLOR = 4
DISPLAY"      OTHERWISE, EXTREME ESTIMATING ERRORS CAN BE
EXPECTED.\"
COLOR = 0
FIND CONTINUE
RESET CONTINUE;

RULE          OUTLIER_REASON_2
IF TRIGGER_OUTLIER_REASON = 2
THEN          OUTLIER_REASON = 2
DISPLAY"The data point {POPPER_OUTLIER_POINT} is believed to
be\"
DISPLAY"an outlier due to an ommitted variable.\"
FIND OMMITTED_VARIABLE
DISPLAY"The ommitted variable {OMMITTED_VARIABLE} is affecting
the data point\"
DISPLAY"{POPPER_OUTLIER_POINT} and that effect is not being
accounted for.\"
COLOR = 4
DISPLAY"EXTREME ESTIMATING ERRORS CAN BE EXPECTED WITH AN

```

OMMITTED VARIABLE."

COLOR = 0

DISPLAY""

FIND CONTINUE

RESET CONTINUE;

RULE OUTLIER\_REASON\_3

IF TRIGGER\_OUTLIER\_REASON = 3

THEN OUTLIER\_REASON = 3

DISPLAY"The data point {POPPER\_OUTLIER\_POINT} is an outlier because it is an"

DISPLAY"anomaly or strange data point. If the point cannot be adjusted,"

DISPLAY"it is acceptable to throw out {POPPER\_OUTLIER\_POINT}. If this point"

DISPLAY" was to be adjusted, it can remain in the data set."

DISPLAY""

FIND CONTINUE

RESET CONTINUE;

RULE OUTLIER\_REASON\_4

IF TRIGGER\_OUTLIER\_REASON = 4

THEN OUTLIER\_REASON = 4

DISPLAY" The data point {POPPER\_OUTLIER\_POINT} is an outlier due to"

DISPLAY" measurement errors. There is nothing you can do about this. "

COLOR = 4

DISPLAY"THIS WILL INCREASE THE POTENTIAL FOR ESTIMATING ERROR IN THIS EQUATION."

COLOR = 0

DISPLAY""

FIND CONTINUE

RESET CONTINUE;

RULE OUTLIER\_REASON\_5

IF TRIGGER\_OUTLIER\_REASON = 5

THEN OUTLIER\_REASON = 5

DISPLAY"The data point {POPPER\_OUTLIER\_POINT} is an outlier due to overriding"

DISPLAY"accounting irregularities. There is nothing you can do about this."

COLOR = 4

DISPLAY"THIS WILL INCREASE THE POTENTIAL FOR ESTIMATING ERROR IN THIS EQUATION."

COLOR = 0

DISPLAY""

FIND CONTINUE

RESET CONTINUE;

RULE OUTLIER\_REASON\_6

IF TRIGGER\_OUTLIER\_REASON = 6

```

THEN      OUTLIER_REASON = 6
DISPLAY"The data point {POPPER_OUTLIER_POINT} is an outlier
due but"
DISPLAY"no information is available to determine why."
COLOR = 4
DISPLAY"THIS MAY INCREASE THE POTENTIAL FOR ESTIMATING ERROR
IN THIS EQUATION."
COLOR = 0
DISPLAY""
FIND CONTINUE
RESET CONTINUE;

RULE      LEVERAGE_REASON
IF TRIGGER_LEVERAGE_REASON = YES
THEN      LEVERAGE_REASON = YES
DISPLAY"{POPPER_HIGHEST_LEVERAGE_VALUE} is not a legitimate
member of the population."
COLOR = 4
DISPLAY"THIS MAY INCREASE THE POTENTIAL FOR ESTIMATING ERROR."

COLOR = 0
ELSE      LEVERAGE_REASON = NO
DISPLAY"{POPPER_HIGHEST_LEVERAGE_VALUE} is an acceptable
member of the population.";

RULE NORMALITY
IF TRIGGER_NORMALITY = NEAR_LINEAR
THEN      NORMALITY = YES
DISPLAY"The normality plot is nearly linear. This suggests
that the residuals"
DISPLAY"are normally distributed. This reinforces the
assumptions of linear"
DISPLAY"regression."
"
ELSE      NORMALITY = NO
DISPLAY"The normality is plot is not considered near linear.
This suggests"
DISPLAY"      that the residuals are not normally
distributed."
DISPLAY"
      There are usually three reasons for
NON_NORMALITY.
      1. The model is misspecified
      2. The model is misidentified
      3. The model is being influenced by outliers"
DISPLAY""
COLOR = 4
DISPLAY" THIS CHALLENGES THE ASSUMPTIONS OF LINEAR
REGRESSION AND INTRODUCES"
DISPLAY"      AN INCREASED POTENTIAL FOR ESTIMATING
ERROR
"
COLOR = 0;

```

```

RULE HETEROSCEDASTICITY
IF TRIGGER_HETEROSCEDASTICITY = 2
THEN      HETEROSCEDASTICITY = NO
DISPLAY"  The property of HOMOSCEDASTICITY hold in this
data set."
ELSE      HETEROSCEDASTICITY = YES
DISPLAY"The property of HETEROSCEDASTICITY has been identified
in this data"
DISPLAY"set.  If this was accounted for with a LOG Y
transformation, the"
DISPLAY"effect is diminished.  If no adjustment was made to
the data set,"
DISPLAY""
COLOR = 4
DISPLAY"THIS CHALLENGES THE ASSUMPTIONS OF LINEAR REGRESSION
AND INTRODUCES"
DISPLAY"      AN INCREASED POTENTIAL FOR ESTIMATING ERROR"
COLOR = 0;

```

```

RULE AUTOCORRELATION
IF TRIGGER_AUTOCORRELATION = NO
THEN      AUTOCORRELATION = NO
DISPLAY"The property of AUTOCORRELATION does not apply in this
data set."
ELSE      AUTOCORRELATION = YES
DISPLAY"  The property of AUTOCORRELATION may be a factor in
this data"
DISPLAY"set.  The DURBIN-WATSON test can be used to identify
AUTOCORRELATION."
DISPLAY"  If AUTOCORRELATION is found to be a problem in the
data set,"
DISPLAY"and it is not corrected, THE POTENTIAL FOR ESTIMATING
ERROR INCREASE.";

```

```

!STATEMENT BLOCK-----
SPECIFICATION-----

```

```

ASK PERCENT:"
Usually, different degrees of accuracy are acceptable at
different phases
of production (less accurate during R & D, more accurate
during production)
What kind of residual values (%) do you expect in this phase
of production?";
CHOICES PERCENT:10%,15%,20%,25%;

```

```

ASK      TRIGGER_OUTLIERS_WRT_Y:
"      Look at the residuals for OBSERVATION.
      Do you have data points with residuals greater than +/-
{PERCENT}?"
CHOICES TRIGGER_OUTLIERS_WRT_Y: YES,NO;

```

ASK SWITCH\_OUTLIERS\_WRT\_Y:  
"List the data point and associated variable with unacceptable residuals.

ENTER A ? AS YOUR LAST ENTRY";

ASK TRIGGER\_OUTLIER\_REASON:  
"Why is this data point coming up as an extreme point?

There are five possibilities:

1. The model is misspecified.
2. Ommitted Variable-the influence of another variable is causing this point's cost to be excessively large or small and that variable is not in the model.
3. This data point is an anomaly or strange data point due possibly to historical perturbations.
4. This data point has an overriding measurement error present.
5. This data point has overriding accounting irregularities.
6. No information available

CHOICES TRIGGER\_OUTLIER\_REASON:1,2,3,4,5,6;

ASK OMMITTED\_VARIABLE:  
"What variable do you suspect was ommitted from the model?";

ASK SWITCH\_HIGHEST\_LEVERAGE\_VALUE:  
"Using either criteria, what DATA POINTS AND ASSOCIATED VARIABLES have unacceptable leverage value?  
(If no variable has a large leverage value according to the criteria, enter the data point with the largest leverage value)  
ENTER A ? AS YOUR LAST ENTRY";

ASK TRIGGER\_LEVERAGE\_REASON:  
"Did you expect this point to have a high leverage value based on earlier analysis of relevant ranges and outliers. If so, this confirms expectations.  
If not, this may indicate that the combination of variables in the equation puts this point in question of being a legitimate member of the population.

Is there any reason to believe that this observation is different from what you are trying to estimate?);

CHOICES TRIGGER\_LEVERAGE\_REASON:YES,NO;

ASK TRIGGER\_NORMALITY:

"Is the normality plot nearly linear at 45 degrees or does the plot depart

substantially from linear?

";

CHOICES TRIGGER\_NORMALITY:NEAR\_LINEAR,NOT\_LINEAR;

ASK TRIGGER\_HETEROSCEDASTICITY:

"Did the model documentation indicate HETEROSCEDASTICITY was a problem?

Choose the example best describing the data points provided with the model:

1. Estimating errors are larger in more expensive systems than for less expensive systems.

2. Estimating errors are approximately the same for all systems";

CHOICES TRIGGER\_HETEROSCEDASTICITY:1,2;

ASK TRIGGER\_AUTOCORRELATION:

"Is the data for this model time-series data or data containing constant lag?";

CHOICES TRIGGER\_AUTOCORRELATION:YES,NO;

ASK CONTINUE:" TO CONTINUE";

CHOICES CONTINUE:\_\_\_\_PRESS\_RETURN\_\_\_\_;

PLURAL: HOLDER\_OUTLIER\_POINT,HOLDER\_HIGHEST\_LEVERAGE\_VALUE;

AUTOQUERY;

SPECCON2.KBS

```
!ACTION BLOCK-----SPECON2-----
ENDOFF;
EXECUTE;
RUNTIME;
BKCOLOR = 3;
```

ACTIONS

LOADFACTS EQUATION

```
LOOP_SPEC= 1
WHILETRUE LOOP_SPEC <= (COUNT_EQUATION_VARIABLES)
THEN
  LOOP_SPEC = (LOOP_SPEC +1)
  GET ALL,EQUATION,ALL
    Equation_Driver    = (Driver)
    Equation_Sign       = (Sign)
    Equation_Derivel   = (Derivel)
    Equation_Derive2    = (Derive2)
  RESET DRIVER
  RESET SIGN
  RESET DERIVE1
  RESET DERIVE2
  GET Equation_Driver = (Driver),YAKBASE,ALL
  RESET EXPERT_DRIVER
  RESET EXPERT_SIGN
  RESET EXPERT_DERIVE1
  RESET EXPERT_DERIVE2
    Expert_Driver      = (Driver)
    Expert_Sign         = (Sign)
    Expert_Derivel     = (Derivel)
    Expert_Derive2     = (Derive2)
  COLOR = 0
  DISPLAY"-----
  -----"
  COLOR = 4
  DISPLAY"CONCLUSION CONCERNING THE DERIVATIVES OF THE EQUATION
  VARIABLES"
  COLOR = 0
  DISPLAY"-----
  -----"
  COLOR = 4
  DISPLAY"CURRENT VARIABLE = {DRIVER}"
  COLOR =0
  DISPLAY""
  DISPLAY"The expert database has classified this variable as
  follows:
  The sign in the equation should be --->
  {Expert_Sign}
  The sign of the first derivative should be --->
  {Expert_Derivel}
```



```

The sign of the second derivative should be --->
{Expert_Derive2}"
DISPLAY"The equation classifies it as:
Sign in the equation ---> {Equation_Sign}
First Derivative sign is ---> {Equation_Derive1}
Second Derivative sign is ---> {Equation_Derive2}
"

FIND MATCH_ALL
RESET MATCH_ALL
DISPLAY""
FIND CONTINUE
RESET CONTINUE

CLS
END

COLOR = 0
DISPLAY"-----"
-----"
COLOR = 4
DISPLAY"CONCLUSION CONCERNING SPECIFICATION LOGIC"
COLOR = 0
DISPLAY"-----"
-----"
DISPLAY""
FIND LOGIC
DISPLAY""
FIND CONTINUE
RESET CONTINUE;

!-----RULES BLOCK-----

RULE ALL_MATCH
IF      Equation_SIGN      =(Expert_SIGN) AND
      Equation_DERIVE1     =(Expert_DERIVE1) AND
      Equation_DERIVE2     =(Expert_DERIVE2)
THEN
    MATCH_ALL = YES
DISPLAY""
DISPLAY"There is no conflict with the specification of this
variable."
DISPLAY"";

RULE NONE_TO_MATCH
IF      Expert_SIGN =? AND
      Expert_DERIVE1 =? AND
      Expert_DERIVE2 =?
THEN
    MATCH_ALL = NONE
DISPLAY""
DISPLAY"There is no information to compare the specification
of this variable."
DISPLAY"";

```

```

RULE SIGN_Mismatch
IF      Equation_SIGN      <>(Expert_SIGN)  AND
      Equation_DERIVE1    = (Expert_DERIVE1)  AND
      Equation_DERIVE2    = (Expert_DERIVE2)
THEN
    MATCH_ALL = NO
    DISPLAY""
    DISPLAY"The expert database and the equation variable are in
    conflict"
    DISPLAY"concerning the SIGN of the variable specification."
    DISPLAY""
    COLOR = 4
    DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
    ESTIMATING ERROR"
    COLOR = 0;

```

```

RULE DERIVE1_Mismatch
IF      Equation_SIGN      = (Expert_SIGN)  AND
      Equation_DERIVE1    <>(Expert_DERIVE1)  AND
      Equation_DERIVE2    = (Expert_DERIVE2)
THEN
    MATCH_ALL = NO
    DISPLAY""
    DISPLAY"The expert database and the equation variable are in
    conflict"
    DISPLAY"concerning the SIGN of the FIRST DERIVATIVE."
    DISPLAY""
    COLOR = 4
    DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
    ESTIMATING ERROR"
    COLOR = 0;

```

```

RULE DERIVE2_Mismatch
IF      Equation_SIGN      = (Expert_SIGN)  AND
      Equation_DERIVE1    = (Expert_DERIVE1)  AND
      Equation_DERIVE2    <>(Expert_DERIVE2)
THEN
    MATCH_ALL = NO
    DISPLAY""
    DISPLAY"The expert databzse and the equation variable are in
    conflict"
    DISPLAY"concerning the SIGN of the SECOND DERIVATIVE."
    DISPLAY""
    COLOR = 4
    DISPLAY"THIS IS CAUSE FOR SOME CAUTION AND MAY INCREASE
    ESTIMATING ERROR"
    COLOR : 0;

```

```

RULE SIGN_AND_DERIVE1_Mismatch
IF      Equation_SIGN      <>(Expert_SIGN)  AND
      Equation_DERIVE1    <>(Expert_DERIVE1)  AND
      Equation_DERIVE2    = (Expert_DERIVE2)

```

```

THEN
    MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"
DISPLAY"concerning the SIGN of the variable specification and"
DISPLAY"the sign of the FIRST DERIVATIVE."
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

RULE SIGN_and_DERIVE2_Mismatch
IF      Equation_SIGN      <>(Expert_SIGN)  AND
        Equation_DERIVE1   = (Expert_DERIVE1)  AND
        Equation_DERIVE2   <>(Expert_DERIVE2)
THEN
    MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"
DISPLAY"concerning the SIGN of the variable specification and"
DISPLAY"the sign of the SECOND DERIVATIVE."
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

RULE DERIVE1_AND_DERIVE2_Mismatch
IF      Equation_SIGN      = (Expert_SIGN)  AND
        Equation_DERIVE1   <>(Expert_DERIVE1)  AND
        Equation_DERIVE2   <>(Expert_DERIVE2)
THEN
    MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"
DISPLAY"concerning the sign of the FIST DERIVATIVE and"
DISPLAY"the sign of the SECOND DERIVATIVE."
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

RULE ALL_Mismatch
IF      Equation_SIGN      <>(Expert_SIGN)      AND
        Equation_DERIVE1   <>(Expert_DERIVE1)      AND
        Equation_DERIVE2   <>(Expert_DERIVE2)
THEN
    MATCH_ALL = NO
DISPLAY""
DISPLAY"The expert database and the equation variable are in
conflict"

```

```

DISPLAY"concerning the SIGN of the variable specification and"
DISPLAY"the sign of the FIRST AND SECOND DERIVATIVE."
COLOR = 4
DISPLAY"THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE
ESTIMATING ERROR"
COLOR = 0;

```

```

RULE LOGIC_1
IF  TRIGGER_LOGIC = 1
THEN  LOGIC = OK
DISPLAY"The logic was set out and followed to get the model
specification."
DISPLAY"The specification can be looked at without risk.";

```

```

RULE LOGIC_2
IF  TRIGGER_LOGIC = 2
THEN  LOGIC = DEVELOP_THEN_RATIONALIZE
DISPLAY"The model was rationalized after it was developed.
This is less than"
DISPLAY"adequate. Unless you can go back to a logic first
approach, you are"
DISPLAY"accepting large amounts of uncertainty and may not be
capturing cost.";

```

```

RULE LOGIC_3
IF  TRIGGER_LOGIC = 3
THEN  LOGIC = GREAT_STATISTICS
DISPLAY"The model was based solely on goodness of fit. This is
less than"
DISPLAY"adequate. Unless you can go back to a logic first
approach, you are"
DISPLAY"accepting large amounts of uncertainty and may not be
capturing cost.";

```

!-----STATEMENT BLOCK-----SPEC CONCLUSIONS

```

ASK CONTINUE:"          TO CONTINUE";
CHOICES CONTINUE:____PRESS_CONTINUE____;

```

```

ASK      TRIGGER_LOGIC:
"The specification logic is the reasoning behind the form the
estimating
equation finally takes. There are different paths that can be
taken to
justify the equation. Which one is the case with this cost
model?

```

1. Start with a set out logic and followed that to get the model equation.
2. Develop the model then rationalize the logic.
3. Just use goodness of fit (go for great statistics).

";  
CHOICES TRIGGER\_LOGIC: 1,2,3;

## ANALYSIS.KBS

```
!ACTION BLOCK-----ANALYSIS AND
EVALUATION-----
ENDOFF;
EXECUTE;
RUNTIME;
BKCOLOR = 3;
```

### ACTIONS

#### LOADFACTS SIGLEVEL

```
COLOR = 4
DISPLAY"-----PART 10  BASIC STATISTICS
-----"
COLOR = 0
DISPLAY""
DISPLAY"This section asks you to look at the basic statistics
of the equation."
DISPLAY"Keep in mind the criteria chosen in Section 2 for
equation acceptance:"
COLOR = 4
DISPLAY"Normal  Acceptance Criteria = {ACCEPTANCE_NORMAL}
percent"
DISPLAY"Unusual Acceptable Criteria = {ACCEPTANCE_EXCEPTIONS}
percent."
COLOR = 0
DISPLAY""
FIND F_TEST
DISPLAY"F-TEST STATISTIC = {F_TEST}"
DISPLAY""
FIND T_TEST
DISPLAY"T-TEST STATISTIC = {T_TEST}"
DISPLAY""
FIND R_SQUARED
DISPLAY"R-SQUARED VALUE  = {R_SQUARED}"
DISPLAY""
FIND CV
DISPLAY"CV VALUE          = {CV}"
DISPLAY""
DISPLAY"ENSURE ALL THESE STATISTICS MEET ACCEPTABLE VALUES
FOR"
DISPLAY"THE USE OF THIS EQUATION"
DISPLAY""
DISPLAY"PRESS ANY KEY TO CONTINUE"
DISPLAY""
DISPLAY"END OF PROGRAM~";
```

```
!RULES BLOCK-----
```

```
!STATEMENT BLOCK-----PART 10
```

ASK F\_TEST: "What is the value for the F-test statistic?";  
ASK T\_TEST: "What is the value for the T-Test statistic?";  
ASK R\_SQUARED: "What is the value of R-Squared?";  
ASK CV: "What is the value for the Coefficient of Variation(CV) ?";

## Appendix C: Knowledge Summary

NOTE: The knowledge summary will follow the same order the files are presented in above in Appendix B. This Appendix contains what is considered to be the accumulation of facts compiled as a result of this expert system. Each fact is indicated by a standard five space indentation.

INTRO.KBS        No Knowledge

HYPER.KBS        No Knowledge

YAKBASE.TXT

The validity of a cost model is based on the following question: "How Well Does The Model Predict What We Want To Predict ?" This is the ultimate question that the analyst will have to answer when evaluating the usefulness of a cost model.

Assumptions for the use of this model are a properly completed Statement of Work, a Request for Proposal that forces adherence to the Sow, and availability of all data and statistics. The first of these assumptions assures the second and third.

A properly completed Sow should request the following information:

1. A list of all the Factors\_that\_influence\_cost
2. A list of all the Key\_cost\_drivers
3. A discussion of Identification\_of\_cost\_drivers
4. A discussion of Specification\_logic
5. A list of all the Raw\_data
6. A discussion of Model\_Properties\_and\_Characteristics
7. A discussion of Outliers
8. A discussion of Ommitted\_Variables
9. A discussion of Heteroscedasticity
10. A discussion of Normality\_Of\_Residuals
11. A discussion of Autocorrelation

A list should be included in the cost model documentation, of all factors that could influence the cost of the population. Each factor identified should be captured by a cost driver in the model.

The key cost drivers are variables which have a specific behavior with respect to cost. These cost drivers are said to "capture" the change in cost.

From a list key cost drivers, an equation is built using



the least squared best fit method. This method may indicated certain key cost drivers better explain the variation in cost than other key cost drivers. The cost analyst must then assess which variables to include and exclude from the equation.

The specification logic deals with the exact form the variable assumes in the cost model equation. Variables may be transformed to indicate different cost behavior. Hence, a premeditated logic should be included explaining the form of a cost driver in the equation.

This includes

- a. The ranges expected
- b. An explanation of cost behavior

The data used to create the cost model should be included with the cost model itself. The raw data must be listed with background information.

This includes

- a. An explanation of any adjustments  
(for example: inflation procedures)
- b. Adjusted data listings
- c. Identification of programs for each data point
  1. Including a brief history
  2. Any events that may have impacted cost of the program  
(for example: A Labor Strike)
  3. Description of the accounting system

The Elements of Model properties and characteristics includes

- a. Model behavior over the range of the data
- b. Significance levels expected for
  1. F-Tests
  2. T-Tests
  3. R-Square
  4. Coefficient of Variation

A discussion of outliers with respect to the independent and the dependent variables must be included. Several methods are available to quantify a data point as an outlier. These will be covered in the cost model.

If a variable is omitted, the cost model may not be capable of capturing some cost. This increases the potential of a bad estimate. Some possible reasons for omitting variables should be considered.

These include

- a. Considerations of lack of data
- b. Collinearity discussions
  1. Among model variables
  2. Among omitted variables
- c. Statistical insignificance discussions

The condition of the error variance not being constant

over all cases is called heteroscedasticity, in contrast to the condition of equal error variance, called homoscedasticity.

(Reference: Applied Linear Regression Models by John Neter - page 423)

Normality plots of the residuals are plots which put each residual against its expected value when the distribution is normal. A plot that is nearly linear suggests agreement with normality, whereas a plot that departs substantially from linearity suggests that the error distribution is not normal. (Reference: Applied Linear Regression Models by John Neter - page 125)

One of the assumptions of basic regression models is that the random error terms are either uncorrelated random variables or independent normal random variables. In some applications, regression involves time series data. For such data, the assumption of uncorrelated or independent error terms is often not appropriate; rather, the error terms are frequently correlated positively over time. Error terms correlated over time are said to be autocorrelated or serially correlated.

(REFERENCE: Applied Linear Regression Models by John Neter - page 484)

The Statement Of Work or SOW is a section of a Request for Proposal. This section specifies what tasks are required for the proper completion of a contract. The contractor is expected to price these tasks and respond to the Request for Proposal with a package explaining his method of accomplishing the items set out in the SOW.

The Request for Proposal is a way the government can solicit priced bids for work. These bids can then be evaluated and a choice can be made based on the most cost effective option.

#### TYPELIST.KBS

The variables that are listed from the database built by the expert and by the cost model should be recorded. The variables that are listed by the expert database and not included in the cost model list of variables should be recorded. The list of variables contained only in the cost model list of variables to be considered should be recorded.

COPYONE.BAT No Knowledge

SIGLEVEL.KBS

You must decide the acceptance levels for model statistics you will use to determine if the equation is significant. The level of acceptance is related to the significance level. The significance level is the Type I error probability.

LEVEL OF ACCEPTANCE = 1 - Type I error probability  
Some equations can be accepted with lower statistics if the included variables are deemed crucial to the model. You should record two different levels of acceptance.

#### EQUATION.KBS

THIS MODEL ASSUMES EACH VARIABLE APPEARS IN THE EQUATION ONLY ONCE.

The equation variables should be picked from the list of cost drivers that is the combination of the cost drivers identified by the experts and the cost drivers that you accept listed only by the contractor.

#### COPYTWO.BAT No Knowledge

#### SETSIZE.KBS

You need to determine how many data points are in the data set provided.

#### OVERLOOK.KBS

You need to determine if there are other systems or data points that could have been included in the data set.

#### SOURRAW.KBS

For each variable in the equation, there is the question of data integrity. For example, "Did the accounting systems provide cost information for these data points from similar systems using acceptable accounting methods?" or "Did the accounting systems provide cost information for these data points were obtained from different systems or by use of unacceptable accounting principles. This is a difficult question to answer, but look at what information you can and attempt to make a determination of confidence in the data for each variable.

For most models, the raw data should be required. Of course, if you did not get the raw data, you will not be able to validate any changes made to it.

When data is adjusted, the method must be obvious and acceptable. If you did not get the data, you can make no assumptions here.

Inflation indices are a common data adjustment. In this case, indices must be provided and applied using an acceptable procedure. Another common adjustment is for differences in quantity. IF THIS IS NOT THE CASE, THE PROBABILITY FOR ESTIMATING ERROR MAY BE GREATER THAN NORMAL.

#### RELEVANT.KBS

The question of relevant range of the data must be looked at for each cost driver. The relevant range is usually determined by the endpoints of the data for each cost driver but not always. Outliers may mislead you into believing the relevant range extends further than is actually the case. Also, consider that you can extend past the endpoints to some degree. This depends on the how much confidence you have that the true function will not deviate much from your extrapolation.

#### HOMOGENE.KBS

This section applies only when you are applying this model to a specific system and you have obtained the necessary data. Here we look a little bit closer at the range of the data. Each data point has one value for each equation variable. The data point you want to estimate also has a value for each equation variable or COST DRIVER (CD).

If the top and bottom of the range for the old system is equal to the value for the new system THEN THE VALUES FOR ALL THE OLD SYSTEMS AS WELL AS THE NEW SYSTEM IS the same. No cost driver is required in this case - the variable does not capture any change in cost due to a change in the CD because the CD is CONSTANT FOR OLD AND NEW SYSTEMS. It can not show up as a significant cost driver.

If the new system value is between the top and bottom of the range for the old systems, this is the ideal situation. Be aware that although the data point is in the relevant range, it may still vary greatly from the dataset with respect to cost.

If the new system value is not between the top and the bottom of the old system range, it is out of the relevant range! You cannot extend to far past the relevant data range without increasing the potential for estimating error. The further you extend outside the relevant range, the less certainty you have that your equation will hold the functional relationship.

If the old systems have all the same value for this cost driver and the new system is different, you cannot measure the influence of the change because this cost driver is constant for the old systems.

### OUTWRTX.KBS

Potential outliers WITH RESPECT TO the X axis(WRT X) can be identified by looking at a graph of data points for each variable. There are four possibilities:

1. An outlier WRT X manifests itself as an extreme point
2. An outlier WRT X is grouped with other points creating gaps in the data
3. Both cases exist
4. Neither case exists

If you cannot determine the reason the point is an outlier the model can be highly influenced by the extreme point. If it appears to be a legitimate member of the population, we don't know if there is any measurement error or not.

THE EFFECT OF GAPS - When Gaps appear in the data set, the behavior between the data point groupings is uncertain. A masking effect may be taking place which introduces additional potential for estimating errors. If the points you are estimating fall in the Gaps, THE POTENTIAL FOR ESTIMATING ERROR IS EXTREMELY HIGH.

When there are no Gaps or extreme points in the data set a positive indication of model integrity is shown.

COPYTHREE.BAT No Knowledge

### VARINFO.KBS

The rate of change of technology has to do with evolutionary changes in the methods and practices in constructing some system type. For example, Aircraft avionics changed from simple to complex. This change indicates an evolutionary change in technology and may affect other cost drivers. A quantum leap in technology can not be captured by any cost driver and invalidates the model. If the equation was built from data on simple avionics airplanes, and you are estimating a complex avionics aircraft, THE RATE OF CHANGE IN TECHNOLOGY IS A FACTOR. Wooden pencils on the other hand are a different story because the technology is the same as when they were invented.

The model assumes none of the variables change direction in the relevant range. The signs of the first and second derivatives must be constant throughout the estimating range.

Determine if the technology is changing so rapidly in this area as to affect the nature of the other factors and variables

Determine which Factor does each cost driver captures.

Determine what form does each variable takes in the equation.

Determine what sign each variable has in the equation.

Looking at new variables, what is the sign of the first derivative and second derivatives?

Determine which variables are composite variables and which are indicator variables in the equation.

#### LISTCON.KBS

No cost drivers listed by the experts were excluded. This indicates that the contractor considered all the relevant cost drivers.

CAUTION - EXCLUDED COST DRIVERS. The following cost drivers were listed in the expert database but NOT listed in the contractors list:xxx. The contractor should have provided a reason for excluding any variables. The experts who built the database see all the cost drivers as important factors for consideration.

Reasons why a cost driver might be left out of the equation.

1. There is an alternate measure (Ex. two different measures of weight).
2. The information for this variable was not obtainable (not measurable).
3. No data available for this cost driver (measurable but not available).
- This variable or cost driver was statistically insignificant when brought into the equation. This indicated
4. Collinearity became a problem when this variable was brought in.
5. The variable is insignificant when combined with certain other variables.
6. The variable is insignificant due to the effects of an outlier.
7. This sample does not represent the population.
8. No reason given

If it is reason 1 above, this is an acceptable reason, NO increase in risk.

If it is reason 2 above, this is an acceptable reason, but this constrains the model.

If it is reason 3 above, this is an acceptable reason, but this constrains the model.

If it is reason 4 above, alternate models should be developed to explore this variable. Any time variables are omitted from the equation collinearity may be a suspected problem. If the cost driver was omitted from the equation to eliminate collinearity problems, the equation may not be capturing all the factor that the variable represented. If the collinear relationship between the omitted variable and the equation variable(s) holds for the new estimate points, the equation is acceptable. If, on the other hand, the relationship between the omitted variable and the equation variable(s) changes for the new estimating point, the problem will not show up as wide confidence bounds but CONSIDERABLE ESTIMATING ERRORS MAY OCCUR.

If it is reason 5 above, this variable was dropped out due to statistical insignificance. The model may be misspecified or misidentified. This is addressed later in the program. Some risk may be present.

If it is reason 6 above, this variable may have been significant if outliers did not exist hence some amount of cost variation may be lost. INCREASED RISK!

If it is reason 7 above, a bad sample misrepresents the population and increases risk.

If it is reason 8 above, when no reason is given for exclusion, risk is can be extremely high.

#### RELEVCON.KBS

When the new point estimate is in the Relevant Range for all cost drivers, this is a positive indication of model integrity.

When your estimate has values outside of the relevant range, for any cost driver, the model behavior is unpredictable. THIS IS AN INDICATION OF HIGH POTENTIAL ESTIMATING ERROR.

#### OVERCON.KBS

This cost model was based on a population which did not include all the members possible. This decreases the integrity of the regression process if the sample used was not chosen randomly.

This cost model was based on a complete dataset. This is a positive indication of model integrity.

#### FACTRCON.KBS

The experts identify factors that should be represented

by variables in the equation. Each variable captures one of the factors listed above. Ideally, there should be one at least one variable per factor.

If there are no variables, the factor is not addressed. This may be a problem. The factors for this system were picked by experts in the field. If a factor is not represented by a cost driver or equation variable, the equation challenges the database built by the experts in this field. This may lead to an increased potential for estimating error.

With one cost driver, you have to make a judgement call. Does this variable capture all the change in the factor? If you think it does, than this is adequate. If not, the change in this factor is not being totally captured and this may result in an inaccurate estimate of cost.

#### IDENCON.KBS

When a new variable is classified as a composite or indicator variable, the expert's database has no basis from which this variable can be compared. The variable stands as specified by the contractor.

When the expert database and the equation variable are in conflict concerning the INDICATOR variable classification or the COMPOSITE variable classification, THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE ESTIMATING ERROR.

#### SPECCON1.KBS

The specification part of the analysis deals with the form of the cost drivers or variables. Most of the data for this section was input previously.

Leverage values will reveal outliers WRT X which are not extreme points but weird combinations of X and Y. There are several ways to determine acceptable leverage values. Two criteria are:

1. A leverage value of .7 or above is considered high
2. A leverage value greater than two times the mean leverage value.

Normality plots plot each residual against its expected value when the distribution is normal. A plot that is a 45 degree line suggests agreement with normality, whereas a plot that departs substantially from linearity suggests that the error distribution is not normal.

Heteroscedasticity is the condition of the error variance not being constant over all cases, in contrast to the condition of equal error variances called HOMOSCEDASTICITY.



AUTOCORRELATION is the condition of the error terms being correlated over time. For this to be a factor the model, the data needs to have a constant lag (time-series data).

If a point is an outlier due to model misspecification, the only remedy for this is to fix the model, (respecification). OTHERWISE, EXTREME ESTIMATING ERRORS CAN BE EXPECTED.

If the data point is believed to be an outlier due to an omitted variable, it is affecting the data point and that effect is not being accounted for. EXTREME ESTIMATING ERRORS CAN BE EXPECTED WITH AN OMITTED VARIABLE.

If the data point is an outlier because it is an anomaly or strange data point and If the point cannot be adjusted, it is acceptable to throw it out. If this point was to be adjusted, it can remain in the data set.

If the data point is an outlier due to measurement errors. There is nothing you can do about this. THIS WILL INCREASE THE POTENTIAL FOR ESTIMATING ERROR IN THIS EQUATION.

If the data point is an outlier due to overriding accounting irregularities, there is nothing you can do about this. THIS WILL INCREASE THE POTENTIAL FOR ESTIMATING ERROR IN THIS EQUATION.

If the data point is an outlier due but no information is available to determine why, THIS MAY INCREASE THE POTENTIAL FOR ESTIMATING ERROR IN THIS EQUATION.

If the normality plot is nearly linear, this suggests that the residuals are normally distributed. This reinforces the assumptions of linear. If the normality plot is not considered near linear, this suggests that the residuals are not normally distributed. There are usually three reasons for NON\_NORMALITY.

1. The model is misspecified
2. The model is misidentified
3. The model is being influenced by outliers

THIS CHALLENGES THE ASSUMPTIONS OF LINEAR REGRESSION AND INTRODUCES AN INCREASED POTENTIAL FOR ESTIMATING ERROR.

If the property of HETEROSCEDASTICITY has been identified in this data set. and If this was accounted for with a LOG Y transformation, the effect is diminished. If no adjustment was made to the data set, THIS CHALLENGES THE ASSUMPTIONS OF LINEAR REGRESSION AND INTRODUCES AN INCREASED POTENTIAL FOR ESTIMATING ERROR.

When the property of AUTOCORRELATION may be a factor in this data set, the DURBIN-WATSON test can be used to identify AUTOCORRELATION. If AUTOCORRELATION is found to be a problem in the data set, and it is not corrected, THE POTENTIAL FOR ESTIMATING ERROR INCREASE.

#### SPECCON2.KBS

If the signs of the derivatives expected by the experts do not match the signs of the derivatives in the equation, THIS IS CAUSE FOR EXTREME CAUTION AND MAY INCREASE ESTIMATING ERROR.

If the logic was set out and followed to get the model specification, the specification can be looked at without risk.

If the logic was developed then rationalized after it was developed, this is less than adequate. Unless you can go back to a logic first approach, you are accepting large amounts of uncertainty and may not be capturing cost.

If the model was based solely on goodness of fit, this is less than adequate. Unless you can go back to a logic first approach, you are accepting large amounts of uncertainty and may not be capturing cost.

#### ANALYSIS.KBS

ENSURE ALL THESE STATISTICS MEET ACCEPTABLE VALUES FOR THE USE OF THIS EQUATION.

### Bibliography

1. Air Force Systems Command. Cost Analysis Procedures. AFSCM 173-1. Washington: HQ AFSC, 1 April 1972.
2. Akers, Michael D. and others. "Expert Systems for Management Accountants," Management Accounting, 67: 30-34 (March 1986).
3. Anderson, Julie. "AI Escapes the Labs," PC TECH Journal, 6: 9 (May 1988).
4. Arcidiacono, Tom. "Computerized Reasoning," PC TECH Journal, 6: 44-56 (May 1988).
5. Dreyfus, Hubert and Stuart Dreyfus. "Why Computers May Never Think Like People," Technology Review, 61: 42-61 (January 1986).
6. Elmer-DeWitt, Phillip. "How to Clone an Expert: Artificial Intelligence Systems are Coming to Market," Time, 126: 44 (September 1985).
7. Fontana, Maxine and Jordene Zeimetz. "Elements of Expert System Shells," PC TECH Journal, 6: 62-65 (May 1988).
8. Freiling, Mike. "Starting a Knowledge Engineering Project: A Step-by-Step Approach," AI Magazine, 6: 150-164 (Fall 1985).
9. Hazen, Capt Christopher M. How Can Air Force Civil Engineers Use Expert Systems? MS thesis, AFIT/GEM/LSM/88S-9. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1987.
10. Holt, Lieutenant Colonel James R. Class handout in LOGM 592, Artificial Intelligence Applications in Management. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, January 1990.
11. Kumara, S. R. T. and others. "Expert systems in industrial engineering," INT. J. PROD. RES., 24: 1107-1125 (1986).
12. Leonard-Barton, Dorothy and John J. Sviokla. "Putting Expert Systems to Work," Harvard Business Review, 66: 91-98 (March-April 1988).

13. McMain, Capt Steven A. An Expert System For Asset Reconciliation. MS thesis, AFIT/GLM/LSY/87S-46. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, September 1987.
14. Moran, Capt Jerry L. ISA: A Prototype Intelligent Scheduling Assistant for Defense Acquisition Management. MS thesis, AFIT/GE/ENG/88M-9. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1988 (AD-A190 481).
15. Murphy, Richard, Instructor in Cost Analysis. Personal interview. Air Force Institute Of Technology, Wright-Patterson AFB OH, 8 March 1990.
16. -----. Instructor in Cost Analysis. Personal interview. Air Force Institute Of Technology, Wright-Patterson AFB OH, 9 April 1990.
17. -----. Instructor in Cost Analysis. Personal interview. Air Force Institute Of Technology, Wright-Patterson AFB OH, 30 April 1990.
18. -----. Instructor in Cost Analysis. Personal interview. Air Force Institute Of Technology, Wright-Patterson AFB OH, 10 July 1990.
19. -----. Instructor in Cost Analysis. Personal interview. Air Force Institute Of Technology, Wright-Patterson AFB OH, 11 July 1990.
20. Nguyen, Tin A. and others. "Knowledge Base Verification," AI Magazine, 8. 69-75 (Summer 1985).
21. O'Keefe, Robert M. and others. "Validating Expert System Performance," IEEE Expert. 81-90 (Winter 1987).
22. Sheil, Beau. "Thinking about Artificial Intelligence," Harvard Business Review, 65: 91-97 (July-August 1987).
23. Pedersen, Ken. Expert Systems Programming: Practical Techniques for Rule-Based Systems. New York: John Wiley & Sons, 1989.

### Vita

Captain Dimitri M. Yallourakis was born on 19 October 1964 in Newburg, New York. He graduated from Concord High School in Wilmington, Delaware in 1982 and attended the U.S. Air Force Academy, graduating with a Bachelor of Science in Engineering in late May 1986. Upon graduation, he received a regular commission in the USAF and served his first tour at Wright-Patterson AFB, Ohio. He began as a Financial Manager for the Training Systems System Program Office where he managed the budget for the F-15 aircraft simulator. He was then chosen to lead the financial team assigned to the F-16 aircraft simulator until entering the School of Systems and Logistics, Air Force Institute of Technology, in May 1989.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302 and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1990	3. REPORT TYPE AND DATES COVERED Master's Thesis		
4. TITLE AND SUBTITLE AN EXPERT SYSTEM FOR THE EVALUATION OF COST MODELS		5. FUNDING NUMBERS		
6. AUTHOR(S)  Dimitri M. Yallourakis, Captain, USAF				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Air Force Institute of Technology WPAFB OH 45433-6583		8. PERFORMING ORGANIZATION REPORT NUMBER  AFIT/GCA/LSM/90S-10		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The Air Force forecasts the costs of new weapons systems with mathematical cost models. A shortage of cost analysts leads to limited evaluations of the cost models. Frequently, analysts look only at the statistical properties of the regression and not at the logic underlying the theory behind the cost model or equation. This research gathered expert knowledge about the cost model evaluation process and created an expert system of 79 rules to assist analysts in evaluating cost models. The rules were implemented in VP Expert 2.0, an expert system shell. This expert system was verified and validated using the experts opinion with 80 percent accuracy. The expert system increased evaluation accuracy which leads to an improved weapons system budgeting process. <i>Key word: Thesis</i>				
14. SUBJECT TERMS Artificial Intelligence, Cost Analysis, <i>(KR)</i> Expert Systems Financial Management		15. NUMBER OF PAGES 149		
		16. PRICE CODE		
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	